
Factors influencing the usability of collision alerting systems in gliding

Einflussfaktoren auf die Gebrauchstauglichkeit von Kollisionswarnsystemen im Segelflug

Dipl.-Ing. Christoph Georg Santel

Dissertation D17

Darmstadt 2016



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Fachbereich Maschinenbau
Institut für Flugsysteme und Regelungs-
technik

Factors influencing the usability of collision alerting systems in gliding
Einflussfaktoren auf die Gebrauchstauglichkeit von Kollisionswarnsystemen im Segelflug

Vom Fachbereich Maschinenbau
an der Technischen Universität Darmstadt
zur
Erlangung des Grades eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigte

Dissertation

vorgelegt von

Dipl.-Ing. Christoph Georg Santel

aus Köln

Berichterstatter: Prof. Dr.-Ing. Uwe Klingauf
Mitberichterstatter: Prof. Dr. phil. Joachim Vogt

Tag der Einreichung: 26. November 2015
Tag der mündlichen Prüfung: 02. Februar 2016

Darmstadt 2016

D 17

Bitte zitieren Sie dieses Dokument als:

URN: urn:nbn:de:tuda-tuprints-51707

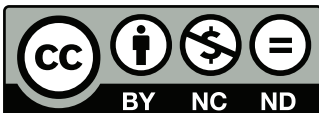
URL: <http://tuprints.ulb.tu-darmstadt.de/5170>

Dieses Dokument wird bereitgestellt von tuprints,

E-Publishing-Service der TU Darmstadt

<http://tuprints.ulb.tu-darmstadt.de>

tuprints@ulb.tu-darmstadt.de



Die Veröffentlichung steht unter folgender Creative Commons Lizenz:

Namensnennung – Nicht kommerziell – Keine Bearbeitung 3.0 Deutschland

<http://creativecommons.org/licenses/by-nc-nd/3.0/de/>

Abstract

The dissertation at hand identifies and analyzes how well glider pilots use low-cost collision alerting systems. While being generally recognized as a commendable tool for helping glider pilots see and avoid other traffic, these systems have been cited as possible contributing factors in several accidents. In literature, no in-depth research on how glider pilots may interpret or misinterpret their indications was found.

At the beginning of this dissertation, a market study of human-machine interfaces for low-cost collision alerting systems is presented. During the study, different human-machine interfaces are taxonomized. The low-complexity and radar-style displays were found to be popular display formats. Also it was discovered that a perspective presentation of traffic has been evaluated for military applications, but not for a gliding context. Thus, a perspective presentation of traffic is proposed. The prototype of a perspective display format for gliding is developed by relying on a user-centered design process. Then, the design features of the low-complexity, radar-style and perspective displays are compared. This results in several hypotheses comparing the usability of the three display formats being postulated.

In order to experimentally evaluate these hypotheses, 137 glider pilots partake in a laboratory experiment. They are presented with traffic information on one of the three display formats installed in a flight simulator. The participants then indicate where they suspect the traffic to be located in the outside world while being exposed to different flight conditions. Performance and subjective satisfaction measurements are recorded during the experiment.

Inferential statistics are used to evaluate the experimental data. The perspective display format results in the most precise estimates of where traffic is located. Generally, errors in estimating traffic position increase as the participants' ownship deviates from straight and level flight. Reaction time does not vary notably between display formats or different flight conditions. Subjective learnability and usability ratings favor the perspective display format over the two other formats analyzed. Overall, the perspective display format exhibits optimized usability in all dimensions when compared to the other two formats.

During the usability analysis, circumstantial evidence arises which suggests that not all participants might interpret the data shown on their display similarly. A probable cause for this may be different knowledge deficits which are experienced between participants. These deficits may result in participants mentally modeling

the collision alerting system incorrectly, thus leading to incorrect coordinate systems for interpreting the traffic information. A method for identifying these mental models is developed. The ensuing analysis reveals that most participants using low-complexity or radar-style display formats incorrectly interpret traffic information in an ownship-fixed fashion. Contrary, most participants working with the perspective display format perform at least some of the required rotations of their personal coordinate systems. The concept of different mental models based on different personal coordinate systems shows potential as an analysis tool for future display designs.

From these findings multiple recommendations are deduced. They are directed at different stakeholders in the gliding community, including glider pilots, aircraft owners and operators, regulatory authorities, glider manufacturers, flight schools, competition rule makers and organizers, as well as designers of collision alerting systems and associated human-machine interfaces. Closing this dissertation, the potential for future human factors research in the gliding community is highlighted.

Kurzfassung

Die vorliegende Dissertation identifiziert und analysiert wie Segelflugpiloten kostengünstige Kollisionswarnsysteme nutzen. Allgemein werden diese Systeme als hilfreiche Werkzeuge wahrgenommen, welche den Piloten bei seiner Aufgabe, Fremdverkehr zu umfliegen, unterstützen. Gelegentlich werden sie auch als potentiell beitragende Faktoren bei Unfällen genannt. In der Literatur lassen sich jedoch nur spärliche Forschungsergebnisse finden, wie zutreffend Segelflugpiloten die Anzeigen dieser Systeme interpretieren.

Den Beginn der Arbeit bildet eine Marktstudie zu Mensch-Maschine-Schnittstellen für kostengünstige Kollisionswarnsysteme. Dabei werden die Mensch-Maschine-Schnittstellen kategorisiert. Der Studie zur Folge sind die Anzeige geringer Komplexität sowie die radarartige Anzeige beliebte Anzeigeformate. Ebenfalls wurde ermittelt, dass eine perspektivische Verkehrsdarstellung bis jetzt nur für militärische Anwendungen untersucht wurde, nicht aber für den Segelflug. Daher wird eine segelflugtaugliche perspektivische Verkehrsdarstellung vorgeschlagen. Dessen Prototyp wird mittels eines nutzerzentrierten Entwurfsprozesses entwickelt. Anschließend werden die Eigenschaften der Anzeige geringer Komplexität, der radarartigen Anzeige und der perspektivischen Anzeige verglichen. Der Vergleich führt zu mehreren Hypothesen bezüglich der Gebrauchstauglichkeit.

Um diese zu prüfen wird eine Laborstudie durchgeführt, an welcher 137 Segelflugpiloten teilnehmen. Jeder Teilnehmer wird in einem Flugsimulator mit Verkehrsinformationen auf einem der drei Anzeigeformate konfrontiert. Während verschiedene Flugzustände durchlaufen werden, geben die Teilnehmer an, wo in der Außenwelt sie den Verkehr vermuten. Messwerte zur Performanz der Probanden und objektivierte Angaben zur subjektiven Zufriedenheit werden während des Versuchs aufgezeichnet.

Anschließend werden die experimentell erhobenen Daten inferenzstatistisch ausgewertet. Die Schätzungen der Verkehrsposition sind mit der perspektivischen Anzeige am präzisesten. Die Positionsschätzung wird fehlerbehafteter, sobald das eigene Luftfahrzeug der Probanden vom horizontalen Geradeausflug abweicht. Bei der Reaktionszeit gibt es keinen nennenswerten Unterschied zwischen den Anzeigeformaten oder diversen Flugzuständen. Zusätzlich wird die perspektivische Anzeige von den Probanden als einfacher erlernbar und besser gebrauchstauglich eingeschätzt. Insgesamt zeichnet sich die perspektivische Anzeige gegenüber den

anderen beiden Formaten durch eine optimierte Gebrauchstauglichkeit in allen Dimensionen aus.

Während der Gebrauchstauglichkeitsanalyse drängt sich der Verdacht auf, dass nicht alle Probanden dasselbe Anzeigeformat gleich interpretieren. Als mögliche Ursache hierfür kommen Wissensdefizite der Probanden in Frage. Diese Wissensdefizite können dazu führen, dass Probanden falsche mentale Modelle der Kollisionswarnsysteme entwickeln. Dies wiederum hätte zur Folge, dass sie inkorrekte persönliche Koordinatensysteme zur Interpretation der Verkehrsinformationen nutzen. Es wird eine Methode entwickelt mit der die unterschiedlichen mentalen Modelle identifiziert werden können. Hierbei zeigt sich, dass die meisten Probanden, welche die Anzeige geringer Komplexität oder die radarartige Anzeige nutzen, die Verkehrsinformationen in einem flugzeugfesten Koordinatensystem interpretieren. Für den Fall der perspektivischen Anzeige führen die meisten Probanden mindestens eine der notwendigen Rotationen ihres persönlichen Koordinatensystems durch. Das Konzept der unterschiedlichen mentalen Modelle, welches auf unterschiedlichen persönlichen Koordinatensystemen beruht, bietet ein passendes Werkzeug für die zukünftige Beurteilung bestehender oder Entwicklung neuartiger Anzeigeformate.

Aus den Ergebnissen der Arbeit werden mehrere Empfehlungen abgeleitet. Diese Empfehlungen richten sich an unterschiedliche Interessengruppen im Segelflug. Dazu zählen unter anderem Segelflugpiloten, Flugzeugeigentümer und -halter, Regulierungsbehörden, Segelflugzeughersteller, Flugausbildungsorganisationen, Ersteller von sportlichen Regelwerken, Veranstalter von Segelflugwettbewerben sowie Entwickler von Kollisionswarnsystemen und deren Mensch-Maschine-Schnittstellen. Abschließend wird aufgezeigt, welches Potential in einer Weiterführung der Forschungsarbeiten zum menschlichen Leistungsvermögen im Segelflug liegt.

Acknowledgments

The dissertation at hand was written during my time as a research associate at Technische Universität Darmstadt's Institute of Flight Systems and Automatic Control. As is the case with most research, it is never truly the work of one person alone. This fact remains true with this dissertation. Colleagues, superiors, students, friends and family and many others have contributed in one extent or another. To all of these people I wish to extend my sincerest gratitude. There are some persons, however, whom I wish to thank personally for their support.

I extend a sincere “thank you!” to the director of the Institute of Flight Systems of Automatic Control, Prof. Dr.-Ing. Uwe Klingauf. Not only did he secure most of the funding for my position and research, but he provided me with the opportunity to freely direct my research to gliding applications. He also acted as my advisor in all doctoral matters. My co-advisor, Prof. Dr. Joachim Vogt, gratefully provided me with access to the resources of his Work and Engineering Psychology research group at Technische Universität Darmstadt as well. Prof. Vogt and his research group always had an open door for all my questions related to psychological matters.

The first year of funding was provided by a doctoral scholarship of the German Research Foundation's Research Training Group 1343 “Topology of Technology”. I thank the representatives of the group for supporting me financially as well as for introducing me to academic disciplines outside of engineering.

Without continued support from my colleagues, performing my research and writing my dissertation would have been a rather dull affair. It was them whom I confronted with my day-to-day ideas. They took the time to discuss my thoughts, and eventually proofread most of this dissertation. Particularly Paul Haiduk, Paul Gerber, Dr.-Ing. Jendrick Westphal and Dr.-Ing. Heike Meinert stood out. They provided me with the everyday discussion and feedback which is so valuable in research.

I am also obliged to several students who contributed by setting up and performing the experiments and studies presented in this dissertation. Verena Schochlow, Simon Mehringskötter, Martin Scheringer and Katrin Heinbücher did a splendid job while working on the experiments, surveys and studies.

Of course, the many glider pilots who participated in this study have my gratitude as well. Unfortunately, but for obvious reasons, they shall remain anonymous. Only with their time and effort was it possible to gather the data presented in this

work. These glider pilots were an incredible group of enthusiasts. Who else would travel through half of Germany at their own expense on a working day, participate in an experiment and be happy about having had a chance to contribute to the sport of gliding?

Support from the gliding community did not only come from individual glider pilots. Representatives from Butterfly Avionics, Garrecht Avionik and Üli's Segelflugbedarf provided me with hardware and support. Through them, I was able to gain insights into the contemporary design processes associated with collision alerting systems and their human-machine interfaces. Within the German Aeroclub, Günter Forneck acted as my point of contact. He enabled me to present my ideas at one of the German Aeroclub's flight instructor meeting and gather input there.

Not all the support required to write a dissertation is technical or professional in nature. My parents Helga and Dr. Hans-Joachim Santel have encouraged me at every step of the way to pursue a career in aviation. Finally, without the love of my significant other, Anja Bott, I am not certain whether I would have found the strength to go through with my dissertation.

Darmstadt, Germany in November of 2015

Christoph Santel



Contents

Nomenclature	xi
1 Introduction	1
1.1 Aim of this thesis	2
1.2 Structure of this thesis	2
2 State of the art	5
2.1 Glider flight operations, collision avoidance and selected technologies	5
2.1.1 Quantifying the risk of midair collisions in gliding	6
2.1.2 Collision avoidance in gliding	8
2.1.3 Collision alerting systems in gliding	9
2.1.3.1 Proliferation of low-cost collision alerting systems .	10
2.1.3.2 Functionality and technical aspects of a quasi-standard low-cost collision alerting system	10
2.1.3.3 Regulatory status in Europe	12
2.1.3.4 Analysis and optimization of low-cost collision alerting systems	13
2.1.3.5 Midair collisions involving gliders with installed collision alerting systems	14
2.1.3.6 Need for research and research questions	15
2.2 Selected aspects of aviation psychology	16
2.2.1 Situation awareness	17
2.2.2 Traffic awareness	18
2.2.3 Mental models	18
2.2.4 Usability	19
2.2.4.1 Measuring effectivity and its influence on traffic awareness	20
2.2.4.2 Measuring efficiency and its influence on workload .	21
2.2.4.3 Measuring user satisfaction	21
2.3 Scientific novelties of the approach presented in this thesis	22

3	Human-machine interfaces of low-cost collision alerting systems in gliding	23
3.1	Commercial-off-the-shelf human-machine interfaces	23
3.1.1	Visual human-machine interfaces	26
3.1.1.1	Dedicated traffic displays	26
3.1.1.2	Non-dedicated displays	35
3.1.2	Auditory human-machine interfaces	36
3.2	Closing the market gap: Developing a perspective traffic display in gliding	36
3.2.1	Conceptual design and prototypical implementation of a perspective traffic display	38
3.2.1.1	Design process	38
3.2.1.2	Display characteristics	41
3.3	Hypotheses	43
4	Determining the influence of display formats on usability	47
4.1	Experiment design	47
4.2	Participants	52
4.3	Equipment	53
4.4	Procedure	57
4.5	Experimental task	60
4.5.1	Feedback during task	61
4.6	Treatment of data and preliminary analysis	62
5	Results on usability and discussion of results	65
5.1	Visual search error (Hypothesis 1)	65
5.1.1	Results	66
5.1.2	Discussion	68
5.2	Explorative analysis: How different ways of showing distance influence the visual search error	69
5.2.1	Results	70
5.2.2	Discussion	71
5.3	Reaction time (Hypothesis 2)	73
5.3.1	Results	73
5.3.2	Discussion	74
5.4	User satisfaction (Hypothesis 3)	75
5.4.1	Results	75
5.4.2	Discussion	76

5.5	Perceived learnability when suppressing familiarity effects	77
5.5.1	Results	78
5.5.2	Discussion	79
5.6	Discussion of technology readiness and overall usability	79
5.7	Criticism of method and potential for future work	80
5.8	Practical implications of results on glider flight operations	82

6	Explorative analysis of mental models and their underlying coordinate systems	85
6.1	Introduction	85
6.2	Method	86
6.2.1	An analytical modeling approach to identifying the mental models used by glider pilots when interacting with collision alerting systems	87
6.2.1.1	Proposing plausible mental models	87
6.2.1.2	Predicting participants' responses	90
6.2.1.3	Assessing the goodness of fit of each mental model	93
6.2.1.4	Selecting the best-fitting mental model for each participant	94
6.2.2	Verification	94
6.2.3	Treatment of data	94
6.3	Results and discussion	94
6.3.1	Frequency of different mental models (Hypothesis 4)	95
6.3.1.1	Results	96
6.3.1.2	Discussion	97
6.3.2	Pilot performance and different mental models (Hypothesis 5)	99
6.3.2.1	Results	99
6.3.2.2	Discussion	102
6.4	Conclusion	104

7	Summary and conclusions	107
7.1	Recommendations	109
7.2	Future scientific work	114

References	117
-------------------	------------

List of figures	131
------------------------	------------

List of tables	133
-----------------------	------------

A Coordinate systems and transformations	135
---	------------

A.1	Coordinate systems	135
A.1.1	Traffic pointer coordinate system	135
A.1.2	Horizon- and track-fixed coordinate system	136
A.1.3	Ownship-fixed coordinate system	137
A.1.4	Horizon- and heading-fixed coordinate system	137
A.2	Coordinate transformations	137
A.2.1	Transformation from horizon- and track-fixed coordinates to traffic pointer coordinates	138
A.2.2	Transformation from ownship-fixed coordinates to traffic pointer coordinates	139
A.2.3	Transformation from horizon- and track-fixed coordinates to horizon- and heading-fixed coordinates	140
A.2.4	Transformation from horizon- and track-fixed coordinates to ownship-fixed coordinates	141
A.3	Describing differences between actual and suspected positions of traffic	141

B Interpretation intervals for effect sizes	143
--	------------

B.1	Cramér's V and Pearson's bivariate r	143
B.2	Pearson's \bar{R}^2	143
B.3	η^2 and generalized η^2	144

C Variations of flight experience and demography with display format	145
---	------------

D Experimentally evaluating the answer sub-task	149
--	------------

D.1	Introduction	149
D.2	Method	150
D.3	Results	151
D.4	Discussion	154

E Verifying the analytical modeling approach to identifying pilots' mental models	155
--	------------

E.1	Introduction	155
E.2	Method	156
E.3	Results and discussion	156
E.4	Conclusion	158

Nomenclature

Scalars	
Notation	Description
$F(M,N)$	Test score of F test with M numerator and N denominator degrees of freedom
$F^*(M,N)$	Test score of Brown and Forsythe’s robust ANOVA with M numerator and N denominator degrees of freedom
\dot{H}	Traffic’s rate of climb
J	Cost function
M	Mean of sample
P	Test score of Page’s trend test
R	Slant distance to traffic
R^S	Slant distance to traffic, as suspected by the pilot
R_H	Horizontal distance to traffic
\bar{R}^2	Ratio of explained variance in multiple regression model, based on Pearson’s adjusted regression coefficient
SD	Standard deviation of sample
$T(df)$	Test score of Student’s t test with df degrees of freedom
U	Test score of Mann-Whitney test
V	Cramér’s V effect size
c	Binary selection switch
f	Model factor
i	Index of mental model
k	Scaling parameter for display symbols and associated labels
k_ϵ	Mapping parameter for mapping elevation information to out-the-window view
k_ρ	Mapping parameter for mapping relative bearing information to out-the-window view
p	Statistic significance
r	Pearson’s bivariate correlation coefficient

Continued on next page

Continued from previous page

Notation	Description
$r_{\text{effect size}}$	Square-rooted ratio of contrast variance to total variance
t_R	Reaction time
w	Weight of cost function summand
ΔH	Traffic's altitude relative to ownship
ΔR	Error in slant distance
Θ	Pitch angle
Λ	Pillai-Bartlett trace
Φ	Bank angle
$\Delta\gamma$	Angular error magnitude between the traffic's actual position and it's position as suspected by the pilot
$\Delta\tilde{\gamma}$	Square-root transformed visual search error
$\Delta\delta$	Error in visual elevation
$\Delta\tau$	Error in visual bearing
$\chi^2(df)$	Test score of χ^2 or Fisher's exact test with df degrees of freedom
α	Predefined significance level of test
δ	Visual elevation
ε	Elevation of traffic, relative to horizontal plane
η^2	η^2 effect size measure
η_G^2	Generalized η^2 effect size measure for repeated measures designs
ν	Drift angle
ρ	Bearing of traffic, relative to ownship's ground track
τ	Visual bearing

Vectors

Notation	Description
\vec{X}^S	Vector from the pilot's ownship to the traffic's position, as suspected by the pilot
\vec{X}^T	Vector from the pilot's ownship to the traffic's actual position
$\vec{x}, \vec{y}, \vec{z}$	Cartesian coordinate axes

Matrices	
Notation	Description
I_3	3×3 identity matrix
M	Matrix of rotation
Superscripts	
Notation	Description
${}^{\prime}, {}^{\prime\prime}, {}^{\prime\prime\prime}, {}^{\prime\prime\prime\prime}, {}^{\prime\prime\prime\prime\prime}$	Intermediate coordinate systems
P	Direction of traffic, as predicted by an mental model
S	Direction of traffic, as suspected by the pilot
Subscripts	
Notation	Description
0	Zero-order term
1	First-order term
ht	Horizon- and track-fixed coordinate system
os	Ownship-fixed coordinate system
p	Perspective display format
r	Radar-style display format
tp	Traffic pointer coordinate system
Θ	Pitch rotation
Φ	Bank rotation
ν	Drift rotation

Mathematical Notation

Notation	Description
\max	Maximum of sample
\min	Minimum of sample
X	Scalar X
\vec{X}	Vector \vec{X}
\mathbf{X}	Matrix \mathbf{X}
\mathbf{X}^{tr}	Transposed Matrix of \mathbf{X}
\dot{X}	Temporal derivative of X
$\angle(\vec{x}, \vec{y})$	Angle between intersecting vectors \vec{x} and \vec{y}
$ X $	Absolute value of X

Acronyms

AHRS	attitude and heading reference system	LED	light-emitting diode
ANOVA	analysis of variance	MAC	midair collision
BFU	German Federal Bureau of Aircraft Accident Investigation	MANOVA	multivariate analysis of variance
CAS	collision alerting system	MM	mental model
COTS	commercial off-the-shelf	OTW	out-the-window
CS	coordinate system	SA	situation awareness
FOV	field of view	SUS	System Usability Scale
HMI	human-machine interface	TA	traffic awareness
		TRL	technology readiness level

1 Introduction

Historically, gliding has been - and still is - a highly popular air sports activity in Germany as well as in other parts of the world. Gliders and motorgliders represent about 51% of the aircraft registered by the German Federal Aviation Office [72]. Similarly, glider pilot licenses represent the majority of flight crew licenses registered there [71]. While gliding is mostly an air sports activity – and therefore “flight for fun, or interest’s sake” [152, p. 18] – it represents an important part of the aviation community.

The gliding community is often a place of leisurely and sportive competitions. Gliding championships and spot landing contests are carried out frequently. This sporting aspect is a source of technological development, ingenuity and research. Erb showed that sport aviation is the economically smallest branch of general aviation in Germany [49, p. 53]. However, especially the gliding community provides high quality technology spill-over to commercial aviation [pp. 163–165]; a fact which is difficult to quantify in monetary units. Thorbeck noted that gliding research is “an [...] incubator for [...] aviation” [142]. He reasoned that it is a source of technological and operational ideas, an environment for conducting basic and applied research and a source of future talents. Hence, one can argue that a functional gliding community is an integral part of the aviation system. Many new technologies, such as composite materials and laminar flow airfoils [20, pp. 105–114] as well as high aspect-ratio wings [pp. 148–158] have been developed for soaring applications, with the know-how then being transferred to commercial applications in powered aviation.

Collision avoidance in gliding is mainly achieved using the see and avoid principle. Wanting to help glider pilots achieve collision avoidance using this principle, another pioneering technology was introduced. Low-cost collision alerting systems (CASSs) have spread swiftly through the gliding community. The groundbreaking introduction of a viable and successful CAS was honored early-on by the gliding community. The design team of today’s quasi-standard FLARM CAS was presented with the OSTIV-Prize in 2006 and with the Prince Alvaro de Orleans Borbon Grant in 2007 [63, p. 1].

Other general aviation operator communities continue to lack a low-cost collision alerting standard. The collision alerting standard mandated for transport-category aircraft [53, AUR.ACAS.1005] results in equipment to be too expensive

and complex to be considered low-cost. Many owners and operators of light aircraft not covered by this mandate may voluntarily install a CAS. Seeing that gliding is an incubator for aviation technologies, it is likely that standardized low-cost CASs will become more popular in other areas of aviation; a process that has already begun.

1.1 Aim of this thesis

Low-cost CASs are an emerging technology in gliding and all of general aviation. By accompanying this emerging technology with scientific research a benefit for the user community of low-cost CASs as well as for the research community is expected. It is the purpose of this thesis to provide scientific input and guidance for future developments in the field of low-cost CASs. Another goal is to sensitize glider pilots as well as flight instructors and training organizations about current shortcomings of CASs. The thesis also documents recent technical developments in the gliding community. Last but not least, it illustrates that human factors research in gliding is necessary and accepted in the community.

1.2 Structure of this thesis

To aid the reader's orientation when reading the thesis at hand, an overview of its structure is given. The core structure, which mainly follows the conventions of human factors research, is also visualized in Figure 1.1.

The first chapter introduces the topic of research and provides initial motivation. It demonstrates that research in the gliding field may provide benefits in terms of technological and scientific spill-over to other areas of aviation.

Chapter 2 begins with a description of contemporary glider flight operations and how collision avoidance is achieved today. An overview of CASs in gliding is given under consideration of technical, legal and safety aspects. Furthermore, questions about the human-machine collaboration between glider pilots and CASs are presented. As these questions regard the field of human factors research, a summary of selected human factors models is given. In chapter 2 the scientific novelties of this thesis, namely a bilateral transfer of research and modeling methods between the fields of aerospace engineering and aviation psychology, are discussed as well.

First findings on the human-machine interfaces (HMIs) used for CASs in gliding are presented in chapter 3. Using a market study, different commercial off-the-shelf HMIs formats are screened. A taxonomy for categorizing these HMIs is given and design characteristics of these HMIs formats are discussed. Two commercially available products, which are in widespread use, are chosen as a baseline

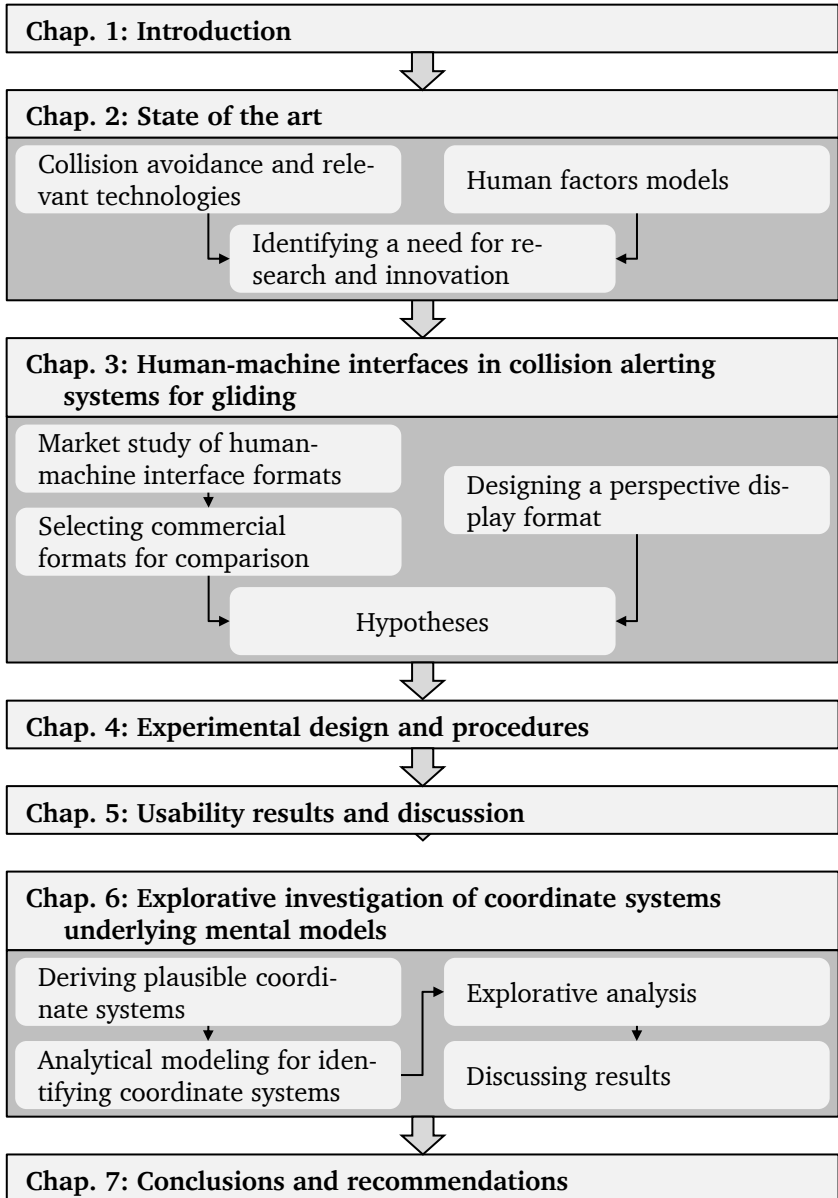


Figure 1.1.: Structure of main part of this thesis. [illustration by author]

for further research. The market study also reveals that a heterocentric perspective display format, currently under development for military applications, is not in use within the gliding community. In this context, a heterocentric perspective display format refers to a graphic representation of the outside world where the display's virtual camera is located at the pilot's ownship position. The camera's attitude, however, is not fixed to the ownship. Instead, it can be fixed to other references, such as the horizon and the ownship's ground track. Seeing a potential benefit in usability by installing this display format in gliders, a perspective display for a quasi-standard low-cost CAS is proposed. With the goal of including potential users in the human-centered design process at an early stage, a prototype display is derived in the later part of this chapter. The chapter concludes with multiple hypotheses on the expected usability benefits of the proposed display format.

In order to test these hypotheses, an experimental setup is presented in chapter 4. The setup relies on exposing participants to different flight conditions in the simulator while interacting with one of three display formats. Subjective and objective usability metrics are determined in the course of the experiment. The chapter not only describes the abstract experimental procedure but also provides details on how these experiments are performed, along with descriptions of the equipment used and the sample of glider pilot participants.

Subjective and objective usability data collected during the experiment is then scrutinized in chapter 5. Inferential statistics are used to test the usability hypotheses of chapter 3. In an attempt to increase the readability of the thesis at hand, these statistical results are interpreted and discussed directly within this chapter. This discussion includes practical implications for glider flight operations with different display formats.

In chapter 6, the data collected during the experiment described in chapter 4 is exploratively analyzed for details on how glider pilots believe the CAS to operate. This approach relies on the assumption that users rely on one of several pre-defined and plausible coordinate systems as a basis for their mental model of the CAS. In an approach known as analytical modeling, those coordinate systems are identified that best describe the performance of each participant during the experiment. Subsequently, the influence of the display format on the mental models of glider pilots is explored and discussed.

The closing chapter of this thesis is used to draw conclusions. Recommendations are extended to the different stakeholders using CASs and their HMIs or taking part in their design. Glider pilots, flight training organizations, avionics manufacturers, aircraft owners and aircraft operators are equally addressed.

2 State of the art

The current chapter provides an overview of contemporary glider flight operations and how collision avoidance is achieved today. Because many glider pilots use low-cost collision alerting systems (CASs), technical and legal aspects of such systems are briefly discussed. Literature is reviewed to determine the state of the art of low-cost CASs regarding their analysis and optimization. As current research does not analyze the human factors implications of these systems for gliding applications, important human factors questions are raised. The second half of this chapter focuses on introducing the human factors tools required to answer these questions. The psychological constructs of situation awareness, traffic awareness, mental models and usability are presented. Ways of quantifying these constructs are also discussed. Chapter 2 closes by summarizing the scientific novelties presented within this thesis. Several research methods are transferred between the fields of gliding research and human factors research.

2.1 Glider flight operations, collision avoidance and selected technologies

As mentioned in chapter 1, gliding is primarily a form of sport aviation. It is usually performed by non-professionals outside of commercial or aerial work operations. Therefore it is a general aviation activity [88, p. 1-5]. However, this does not imply that gliding is unregulated. Training and licensing are required for glider pilots to operate their aircraft.

While flying any heavier than air aircraft, aerodynamic drag of the aircraft dissipates energy [21, section 3.4.1]. Because gliders lack an engine to compensate this energy dissipation, glider pilots must seek updrafts or other meteorological phenomena. Using these phenomena, gliders can remain airborne for extended periods of time. Reichmann differentiated between five meteorological phenomena which are used in gliding to stay aloft [118, pp. 16–65].

Thermal updrafts are caused by solar radiation. Given a sufficiently unstable vertical temperature profile, convective cells develop. The core of these cells consists of rising air currents and can be several hundred meters wide [84, section 4.2]. Glider pilots usually circle in this core area to gain altitude.

Ridge lift is a form of forced convection. It develops whenever wind flows perpendicular to a long obstacle, such as a mountain ridge. The air mass is forced over the obstacle and rising air exists at the windward side of the obstacle.

Mountain lee waves occur on the downwind side of mountain ridges or individual mountains. These waves require a stable vertical temperature profile and adequate wind speeds [118, p. 56]. The updraft regions, which may extend several kilometers laterally, are usually characterized by continuous and laminar vertical wind components.

Inversion waves require strong wind shear near the altitude of a temperature inversion. Their wandering location makes inversion waves difficult to use in practical glider operations.

Dynamic soaring is a means of extracting energy from a wind field with vertical wind gradient [81]. Due to the high load factors required for continuous flight, this method of soaring is rarely used in manned applications.

By far the most frequently used meteorological updraft phenomena in gliding are thermal updrafts and ridge lift. Since both updraft phenomena are extremely limited in their lateral extensions, high local traffic densities of multiple gliders with an increased risk of collisions may occur. Figure 2.1 illustrates high local traffic densities which may result in a thermal updraft during a gliding competition. Due to the ever-present risk of a midair collision (MAC), glider pilots are specifically trained for flights in proximity to other aircraft in these updrafts [75, section 2.5.1; 58, chap. 10, sections “Thermal soaring” and “Ridge and slope soaring”].

2.1.1 Quantifying the risk of midair collisions in gliding

An inherent risk of MACs exists whenever multiple gliders fly in close proximity to one another. The German Federal Bureau of Aircraft Accident Investigation (BFU) registered a total of 1841 accidents involving gliders and motorgliders from the year 2000 to 2013 (see Table 2.1) [73]. Out of these occurrences, a total of 267 involved fatal injuries. MACs involving gliders or motorgliders made up 57 of these accidents. A total of 26 MACs resulted in fatalities.

Performing Fisher’s exact test on Table 2.1 revealed that the number of fatal accidents was higher for MACs than for other accident categories. This was a small, yet non-negligible, effect;¹ test score $\chi^2(1) = 45.92$, significance level $\alpha = .05$, significance $p < .01$, Cramér’s V effect size $V = .158$.

¹ The interpretation of effect sizes is discussed in appendix B.



Figure 2.1.: High local traffic density underneath a cumulus cloud during the 32nd FAI World Gliding Championships of 2012 in Uvalde, TX. [photograph and illustration by author]

Table 2.1.: Accidents involving gliders and motorgliders between 2000 and 2013, as registered by the German Federal Bureau of Aircraft Accident Investigation (BFU).

Severity	accident category	
	midair collisions ^a	other
Fatal	26	241
Non-fatal	31	1543

^a The BFU's database [73] categorized all registered occurrences according to the Civil Aviation Safety Team / International Civil Aviation Organization Common Taxonomy Team's taxonomy [30]. This taxonomy also defined near midair collisions as midair collision events [p. 3].

MACs are rare, representing only 3.1% of all accidents. A similar value of 2.1% was reported by van Doorn and de Voogt in their analysis of United States glider accidents between 2001 and 2005 [42]. However, MACs have a much higher risk of involving fatalities when occurring. While only 13.5% on non-MAC occurrences are fatal, this ratio increases to 45.6% for MACs.

Causes and factors contributing to MACs of powered aircraft are generally well-documented [103; 140; 151] and internationally consistent [10]. In the mentioned literature there is general agreement that most MACs share the following traits.

- The collisions occur in the vicinity of an airport at an altitude below 2000 ft above ground level, and
- the collisions occur during daytime in good visual meteorological conditions.

Also, Shuch suspected that the quantitative risk of being involved in a MAC is principally overestimated by pilots [130]. He also estimated that, as pilot experience increases, the perceived risk of being involved in a MAC approaches the actual risk.

However, the previously mentioned research on characteristics and perception of MACs mostly focused on powered aircraft and did not analyze the peculiarities of glider flight operations. Janke et al. analyzed near-misses and MACs registered by German authorities between 1980 and 2001 [89, section 5.4]. Apparently, near-misses and MACs in Germany are most common between gliders. According to their assessment, 47.3% of German near-misses and MACs involved two gliders or motorgliders that were both either thermaling or ridge soaring.

2.1.2 Collision avoidance in gliding

Gliding is almost exclusively carried out under visual flight rules during daytime in visual meteorological conditions. Air traffic control only provides separation between visual flight rules traffic in Class B airspace [55, SERA.6001]. In all other airspace classes, no separation is provided by air traffic control, and thus, glider pilots must self-separate.

Successful self-separation without the help of onboard tools or air traffic control requires visually acquiring other traffic. This concept is known as the see and avoid principle. Because the traffic is visually acquired by human operators, a human component exists in the air traffic system. Flight experiments conducted by Andrews demonstrated that see and avoid is a non-deterministic process [6]. According to his work, the likelihood of visual detection of another aircraft depended on a multitude of factors, including pilot alertness, target angular size, target contrast and search time available to the pilot. Due to its stochastic nature see and

avoid cannot always guarantee visual target acquisition, which is the prerequisite for evasive maneuvering.

Generally, the lack of visual detection of another aircraft is without consequence, unless the aircraft are on a collision course. In the latter case, consequences can be dramatic, culminating in a MAC. The causes for overlooking traffic may be technical, physiological or psychological. Technical causes can be wings or other airframe components shielding another aircraft from sight [89, p. 30]. Physiological reasons may include low target conspicuity, glare or lack of relative motion. Comprehensive discussion of the aeromedical factors involved with vision processing can be found in literature, such as in works by Gibb, Gray, and Scharff [74, chaps. 1–4], Amendt, Knebel, and Wolff [3] and Campbell and Bagshaw [27, section 3.2]. Psychological causes are always present if an adequate physical stimulus for visual detection of another aircraft exists but this stimulus is not recognized by the pilot.

On a personal basis, the see and avoid principle is executed by the pilot as part of her or his visual scan. The visual scan is the procedure by which the pilot visually surveys the instrument panel as well as the out-the-window view. Guidance is provided to pilots on how to perform these scans. For example, the Federal Aviation Administration [59, section 8-1-6. c.] and International Civil Aviation Organization [87, p. 8] recommended that pilots spend between 67% and 80% of their time visually scanning for traffic in a structured pattern. However, Colvin, Dodhia, and Dismukes challenged the guidance as being impracticable [33]. Also, the guidance is directed at pilots of powered aircraft and not directly applicable to the specifics of glider flight operations. Apel emphasized that glider pilots must maintain a constant lookout, good situation awareness and a high degree of self-control whenever thermaling with other gliders [7, section 2.9.4].

2.1.3 Collision alerting systems in gliding

Miniaturization and declining costs of electronic components in the late 1990s have introduced CASs as a whole new class of pilot assistance systems in the gliding community. Literature revealed no definition for these low-cost CASs. Therefore, the following definition is introduced:

Definition 1: *Low-cost collision alerting systems* are a class of pilot assistance systems aimed at increasing pilot traffic awareness while complying with the technical, operational and economic requirements of sport aviation.

These systems usually attempt to increase the traffic awareness of pilots by guiding their attention to traffic and aiding in the traffic's visual identification. Andrews

showed that a well-designed CAS decreases the time needed to visually acquire traffic with a given likelihood by guiding the pilot's gaze into the direction of the traffic [5, p. 37]. Accident investigators see CASs as commendable and practical tools for avoiding MACs [68, p. 12].

Several CASs developed specifically for gliding and air sports applications have been patented at the turn of the century. Systems based on satellite navigation [139] as well as radio direction-finding [116] have been proposed at the time. They fulfill the definition of a low-cost CAS. While not specifically developed for the needs of the gliding community, several other sensing technologies are used in contemporary low-cost CASs worldwide. These include systems based on transponder signal analysis [67, section 6.2], Automatic Dependent Surveillance-Broadcast [section 7.3.2], and Traffic Information Service-Broadcast [section 7.3.1].

2.1.3.1 Proliferation of low-cost collision alerting systems

While not required under European air law [54], low-cost CASs have become popular in the European gliding community. 135 of 137 glider pilots participating in the experiments presented in this thesis have flight experience with low-cost CASs.

A statement by the Swiss Civil Aviation Safety Officer offered a glimpse into the distribution mechanisms.

The rapid distribution of such [low-cost collision alerting] systems only a few months after their introduction was not accomplished through regulatory measures, but rather on a voluntary basis and as a result of the wish on the part of the involved players to contribute towards the reduction of collision risk. [137, p. 103]

This indicated that social mechanisms were apparently involved in the distribution of low-cost CASs. These mechanisms may include concerns for pilot's safety as well as social pressures exerted onto the owner or operator.

2.1.3.2 Functionality and technical aspects of a quasi-standard low-cost collision alerting system

According to an unpublished survey of 699 German glider pilots by Steinmetz and Gerber, a commercial product derived from a European patent [139] is established as a quasi-standard of low-cost CAS for gliding applications in Germany [134].

This CAS product is known as the FLARM² system. According to the manufacturer “FLARM is designed and built as a non-essential ‘situation awareness only’ unit to only support the pilot, and cannot always provide reliable warnings. In particular, FLARM does not give any guidance on avoiding action” [64, p. 11].

The quasi-standard FLARM CAS was introduced as a commercial product in 2004. It was specifically developed for the requirements of the gliding community [123]. It is a cooperative CAS, requiring all participating aircraft to be equipped with a compatible transceiver box (Figure 2.2). Each transceiver box determines its own position and track velocity using Global Positioning System and barometric measurements. This position and velocity information is then broadcast to other transceivers using a proprietary radio protocol. Each transceiver uses flight phase and threat detection algorithms to assign threat levels to the received traffic [64, pp. 2–3]. This threat information is displayed on a simple human-machine interface (HMI) integrated into the transceiver’s front. The integrated HMI has the capability to show only one traffic aircraft at a time. However, information on all traffic received can be passed to third-party HMI devices through a publicly documented data port [65]. The system’s operating principle is illustrated in Figure 2.3.



Figure 2.2.: FLARM transceiver box of hardware version 3. [photograph by author]

Also, no attitude and heading reference system is incorporated into the FLARM system. While this decreases system complexity along with development and production costs, it has a major influence on HMI design, as will be shown in chapter 3. All traffic information must be derivable from geo-referenced Global Positioning System measurements.

² In some literature *FLARM* is defined as *FLight ALaRM*. However the manufacturer, FLARM Technology GmbH of Baar, Switzerland, does not utilize this definition. Instead, the term *FLARM* is used as a proprietary trade name.

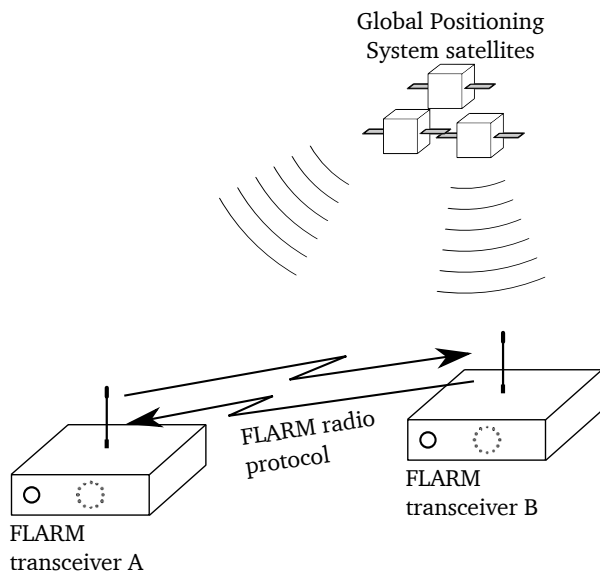


Figure 2.3.: Conceptual sketch of the FLARM system. [illustration by author]

Being a non-certified avionics system, FLARM breaks with several conventions of avionics design [63]. For example, all FLARM transceivers are equipped with the same Global Positioning System receiver model and receiver firmware. Furthermore, the radio protocol through which the transceivers communicate with one another is proprietary and not open to the public. Instead, the manufacturer only permits original FLARM equipment to participate in the system. These breaks in convention are argued to lower the production and sales costs of FLARM transceivers to a level which is acceptable for sports aviation. At least one competitor challenges the legal and safety aspects of the radio protocol's proprietary nature. DSX High Tech participates in the FLARM system with a transceiver unit that is not endorsed by the FLARM manufacturer [43].

2.1.3.3 Regulatory status in Europe

European air law does not require gliders, motorgliders and other light aircraft to be equipped with CASs [54]. However, the European Aviation Safety Agency considers these systems to be standard equipment for gliders and motorgliders [50,

AMC 21.A.303(c)]. By taking this approach, CAS manufacturers are able to circumvent a costly certification process for their products.

Also, self-regulation within the gliding community makes carrying the quasi-standard FLARM system mandatory in some regions. The French Gliding Federation requires FLARM to be installed on all gliders and tow planes operated by its clubs [121]. Furthermore, recommendations were extended by de Boer to the International Gliding Commission to honor the use of low-cost CASs in its contest regulations [18].

2.1.3.4 Analysis and optimization of low-cost collision alerting systems

Janke et al. considered FLARM to be the most promising of four low-cost CASs under development at the time of their study [89, pp. 80–83]. However, the authors did caution that FLARM would need to be evaluated once being more widely established. Since then, several technological improvements and studies of the system have taken place.

Influences of high traffic densities on the FLARM radio protocol were studied by Schuler [127] and Berweger and Schuler [16]. Their work showed that the radio protocol is adequate even for extremely high traffic densities. A German patent suggested that there is potential for improving the antenna hardware of FLARM [40]. Baumgartner and Maeder also proposed improvements to FLARM's flight phase detection algorithms [14]. According to their work, there is potential for improving the algorithms in high-wind conditions.

According to FLARM Technology, the design of HMIs is left to industrial partners [63, p. 3]. These external HMIs rely on being fed with the data stream from each FLARM transceiver [65]. From a human factors perspective, proper HMI design is important since the HMI “acts as a medium between some aspects of the actual information in a system [...] and the operator's perception and awareness of what the system is doing, what needs to be done, and how the system functions (the mental model)” [155, p. 185].

However, literature research revealed no studies or comprehensive guidance material discussing human-machine collaboration for the case of low-cost CASs in gliding. Only trivial guidance, such as the recommendation that CASs “should have audio warnings and direct your eyes to the 'threat'” [57, p. 18], is given. This suspicion was fortified by representatives of companies offering FLARM-compatible HMIs.³ The research available focused on how pilots of commercial aircraft [13; 150] and powered general aviation aircraft [77; 78; 157] interact

³ J. Garrecht, personal communication, August 22, 2011; M. Förderer, personal communication, September 24, 2012

with CASs and their HMIs. Generalizing this knowledge to gliding applications does not seem prudent. The operational demands in glider flight operations are suspected to be much too different compared to aviation operations.

2.1.3.5 Midair collisions involving gliders with installed collision alerting systems

Accident investigators repeatedly recommended equipping general aviation aircraft with CASs to reduce the number of MACs [76; 82]. However, CASs have not been able to prevent all MACs. Multiple collisions between gliders and/or motorgliders equipped with operational CASs have occurred in Germany [92], Finland [95] and the United States [105, NTSB Accident ID CEN12LA553AB]. Neis discussed that not only technical aspects - but also human factors - contributed to this series of MACs [107]. The reports from Knoll et al. [92, p. 8] as well as Laine et al. [95, p. 15] describe insufficient traffic awareness of the pilots involved as factors during these accidents. Knoll et al. also reported a contributing cause. The glider pilots did not adequately interpret CAS signals. The low-cost CASs installed in the involved gliders did not fulfill their design goal of increasing pilots' traffic awareness to a sufficient level.

In order to test whether the annual rate of MACs involving gliders has declined since the introduction of FLARM, the BFU's accident data of Table 2.1 was analyzed further [73]. The annual number of BFU-registered MACs involving gliders is depicted in Figure 2.4. The annual MAC rate of the pre-FLARM years of 2000 through 2006 was compared to the more recent years, when FLARM was established. No exact data on how fast FLARM established itself in the German gliding community was available. However, Scherer first discussed the system in a pilot's magazines in 2006 [123]. It was assumed that FLARM reached its status as a quasi-standard by the end of that year.

Twenty-nine MACs occurred between the years of 2000 and 2006 whereas twenty-eight collisions took place from 2007 to 2013. A one-sided two-sample t test for equal variances⁴ was performed on the data. It provides statistically insignificant results (test score $T(12) = 0.17$, $p = .43$). The data provides no indication that the introduction of FLARM has reduced the annual rate of glider and motorglider MACs.⁵ While not yet statistically traceable, BFU investigators still sus-

⁴ A Levene test of the data's variances provided statistically insignificant results.

⁵ To determine the annual rate of MACs, normalizing data on the annual flight time or number of flights is required. Unfortunately, no reliable data is available at the time of this writing. The European Aviation Safety Agency is currently gathering this information for future safety assessments [51, p. 5; 52, p. 4].

pected that the number of MACs and near-MACs per flight hour have declined due to the introduction of FLARM.⁶

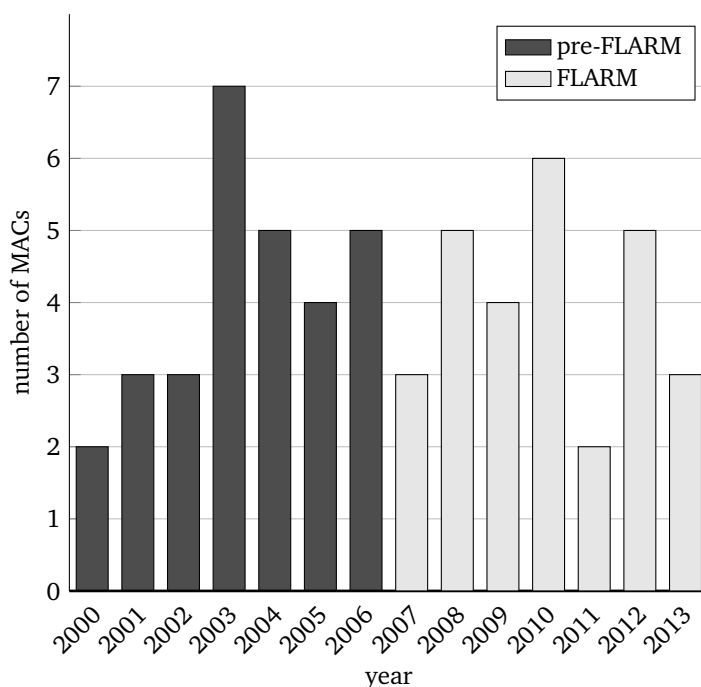


Figure 2.4.: Midair collisions (MACs) involving gliders and motorgliders between 2000 and 2013, as registered by the German Federal Bureau of Aircraft Accident Investigation.

2.1.3.6 Need for research and research questions

Section 2.1.1 showed that MACs involving gliders and motorgliders occur rather infrequently when compared to other accident causes. When MACs occur, their fatality rates are much higher than those of other accident categories. As shown in section 2.1.3, low-cost CASs were introduced into the gliding community with the goal of mitigating the risk associated with MACs. However, these systems provide

⁶ K.-U. Fuchs, personal communication, September 4, 2015

no absolute protection against MACs and the annual rate of MACs in the German gliding community has not dropped notably since the introduction of low-cost CASs (section 2.1.3.5). One factor possibly contributing to the collision of CAS-equipped gliders is inadequate traffic awareness of the flight crews involved, associated with improper comprehension of the HMI signals presented to the pilots. When wishing to prevent this contributing factor from occurring in future accidents, it is necessary to understand how HMIs of CASs in gliding can be misinterpreted. Then non-ambiguous HMIs can be designed. However, guidance material on proper HMI design for the application at hand is lacking in literature. This reveals a scientific need to study the HMIs of low-cost CASs in gliding applications.

Research questions may be:

- Do current HMIs of low-cost CASs adequately assist glider pilots in visually acquiring and avoiding conflicting traffic?
- Do they assist in increasing the pilots' traffic awareness?
- How can current HMIs be improved to increase the usability of low-cost CASs for gliding applications?

All of these questions discuss how low-cost CASs influence the pilot. Seeing that the pilot has to act upon the information presented, the discussion becomes a matter of how low-cost CASs influence the pilot's behavior. The nature of these questions is outside the scope of a purely technical discussion. Instead they belong to the field of aviation psychology and usability research. In the following section, an overview of psychological constructs used to answer these questions is given.

2.2 Selected aspects of aviation psychology

The preceding section shows that collision avoidance in glider flight operations is not solely achieved through technical means. Instead, the pilot must interpret the information shown on the CAS's HMI and determine "what needs to be done" [155, p. 185] to avoid a potential collision. This requires the pilot to have adequate awareness of the traffic nearby as well as an understanding of how the CAS works. How well this awareness is supported by the low-cost CAS is a question of how usable the system is for this task. Adequate usability can be achieved by appropriate design processes.

2.2.1 Situation awareness

A term often mentioned in discussions of pilot assistance systems is the concept of situation awareness (SA) [64, p. 11]. Endsley provided a commonly used definition of the concept. “Situation Awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” [46, p.792]. This definition describes SA as a multi-level process, which was also discussed by Endsley, Bolté, and Jones [48, chap. 2].

1. **Perception** As a first step environmental elements must be perceived by the physical cues they offer to the pilot.
2. **Comprehension** After physical cues are observed, they can be integrated into an understanding of the current state of the elements and compared to their goals and intentions.
3. **Prediction** Based on the comprehension of the current states and their deviation from the goals of the elements, the pilot is capable of estimating the elements’ future states. This is a prediction of element behavior.

Durso and Alexander showed that SA is not an independent construct, but closely intertwined with other concepts of aviation psychology [44]. Any change in SA leads to changes in the pilot’s performance and workload and vice versa. They also argued that SA can be quantified by implicit performance metrics. For example, it is possible to task the pilot with providing her or his estimate of traffic position and define the pilot’s error in relation to the actual traffic position as a SA measure [pp. 225–226]. Also, it is possible to use the pilot’s response times to “the main task of interest” [p. 230] as a workload measure.

Nevertheless, SA as a concept is not free of criticism. Dekker and Hollnagel considered SA to be a folk model in human factors [36], which (a) is prone to focus “on descriptions rather than explanations” [p. 80], (b) is being resistant to falsifications and (c) leads often to overgeneralization. However, Parasuraman, Sheridan, and Wickens particularly rejected Dekker and Hollnagel’s criticism on the lack of falsifiability of SA [115].

Casner cautioned about the role of advanced cockpit systems, such as low-cost CASs, in the general aviation cockpit [29, pp. 604–607]. In many situations, these systems may have detrimental effects on SA, workload and performance. This caution serves as a further motivation for the research presented. It also leads to the question of how the detrimental impact of these systems may be reduced.

2.2.2 Traffic awareness

In the preceding sections, the concept of traffic awareness (TA) was implicitly discussed. It was also an explicit part of Definition 1. Literature treated TA mainly by working descriptions and categorizations, without offering an explicit definition.

Alexander and Wickens described TA to be “one of a general class of situation awareness (SA) measures which assesses the pilot’s accurate and timely understanding of the representation of traffic in 3D space” [2, p. 171]. According to Wickens TA is a subcategory of spatial awareness [153] while Uhlarik and Comerford described it as a surveillance-category component of SA [148, p. 14]. In both cases TA was considered to be a subcategory of SA. Therefore, it is subject to the same influences as SA.

For this thesis an explicit definition of TA is introduced. It is defined in analogy to Endsley’s definition of SA [46].

Definition 2: *Traffic awareness* is the perception of other aircraft on the ground and in the air within a volume of time and space, the comprehension of their current phase of flight and flight state, and projection of their future positions, intentions and maneuvers.

The multi-level structure of SA is retained in this definition of TA. It is illustrated in Figure 2.5. In the case of visual flight rules operations TA should not be derived from the CAS alone, but rather from the out-the-window view.⁷

2.2.3 Mental models

In order to comprehend (level 2 SA/TA) and predict (level 3 SA/TA) how an arbitrary system - such as a CAS - behaves, the pilot must have an idea of how the system functions [47]. This idea is often referred to as being the pilot’s mental model (MM) of the system. Rouse and Morris defined MMs as being “the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states” [122, p. 7].

Such MMs can take on many shapes, from a mere *black box* correlation between input and output to an intricate understanding of the system’s internal processes. The models are also subject to change over time, being influenced by the pilot’s

⁷ In commercial and military aviation, the context of receiving traffic information is different. Traffic is often beyond visual range, and therefore out of sight. In this case TA must be derived from the HMIs alone.

perceptions (level 1 SA/TA). Figure 2.5 illustrates this interaction between TA and the MM. Wickens et al. also pointed out that a pilot's MM of a system is influenced by the HMI of that system [155, pp. 188–189]. Therefore, an influence of the CAS HMI on both, the pilot's MM of the CAS and on TA, is suspected. In addition, Endsley et al. emphasize that incomplete or false MMs can lead to dangerous breakdowns of level 2 and 3 SA/TA [48, pp. 40–41].

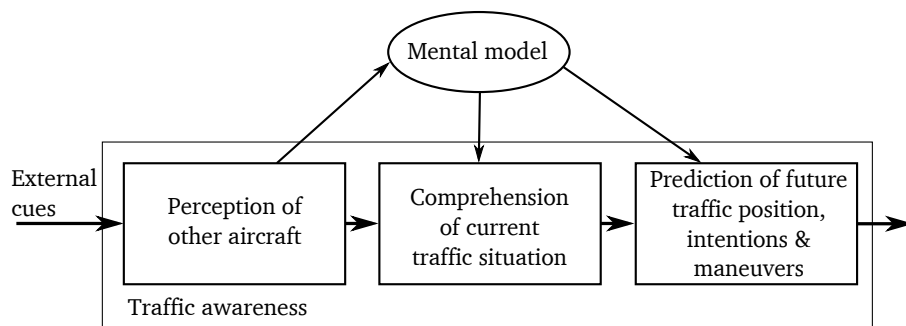


Figure 2.5.: Levels of traffic awareness and their interaction with a mental model. [illustration by author, analog to Endsley [47]]

Rouse and Morris described several methods for identifying MMs [122, pp. 8–17]. One such approach is analytical modeling. It requires making multiple educated guesses about potential model structures that a person may have of a system. Afterwards, experimental data is generated by exposing the person to the system. The proposed models are then compared to the experimental data. The model which fits the experimental data best is selected as being the most-likely MM that the person may have. This approach is comparable to systems identification in engineering. However, analytical modeling does not guarantee that the person's actual MM is detected. If the actual MM has not been proposed by the investigator a-priori, then it cannot be detected [pp. 11–13].

2.2.4 Usability

As per Definition 1, a low-cost CAS must have the design goal of increasing the pilot's TA. It was shown in section 2.2.3 that changes in the HMI of a low-cost CAS may result in different MMs. These different MMs may then affect TA. How well the system and HMI increase the pilot's TA is a question of usability research.

An international standard defined three dimensions of usability (see Figure 2.6) [131, section 5.4].

Effectivity is one of two user objective performance dimensions of usability. It measures the precision and completeness of how a user performs her or his task with using a system.

Efficiency is the second objective performance dimension. It relates the effectivity to the user's effort. Different dimensions of effort are the time spent on a task, as well as cognitive, material and monetary effort.

User satisfaction is a subjective measure of the user's experience while interacting with the system. It reflects whether the user considers the system to be free of deficits. In general, satisfaction regards the user's attitude towards the system.

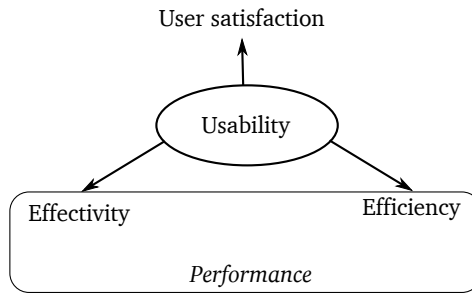


Figure 2.6.: Dimensions of usability. [illustration by author]

Bevan and MacLeod clarified that usability is not only an attribute of a system, but must be seen in context of the task that the user performs with the system [17]. Gliding is a complex task performed by glider pilots while attending to multiple sub-tasks. One of these sub-tasks is the visual identification of nearby traffic. Ideally, it is supported by a CAS.

2.2.4.1 Measuring effectivity and its influence on traffic awareness

In this thesis, effectivity describes how accurately the HMI of a low-cost CAS guides the glider pilot's glance into the direction of the indicated traffic. If the pilot's glance is guided into the direct vicinity of traffic, it should be simple for the pilot to visually identify the traffic. Therefore, TA should be good.

Before consciously moving their glance to the direction of suspected traffic, the pilots must form an idea where the traffic may be located. When \vec{X}^T is the vector

from the pilot's ownship to the traffic's actual location, and \vec{X}^S is the vector to the traffic's location as suspected by the pilot, then the angular error $\Delta\gamma = \angle(\vec{X}^T, \vec{X}^S)$ can be determined. Whenever the angular error is low, then effectivity is high and vice versa.

Durso and Alexander pointed out that the pilot's estimation of traffic position is not only a measure of effectivity but also an implicit performance measure of TA [44, p. 223]. As mentioned above, higher effectivity (lower angular errors) will lead to swifter visual traffic identification (perception; level 1 TA) by the pilot and therefore result in better TA.

2.2.4.2 Measuring efficiency and its influence on workload

In the study at hand, efficiency is considered to be the relationship between effectivity and the glider pilot's workload experienced during the task of locating traffic with the CAS. Reaction times to the task of estimating traffic locations from a CAS HMI are a primary task performance measure of workload [44, p. 230]. Therefore, reaction times influence efficiency. This relationship between effectivity and reaction times is classified as temporal efficiency [17], which will be the focus of this thesis.

Piloting an aircraft is a process placing multi-task demands on the pilot, resulting in a certain level of workload [154]. Popular workload measures in aviation include physiological and subjective self-reporting measures [79]. Because reaction times are sensitive to workload and correlate positively with an increase in workload [109, p. 42-21], temporal efficiency was identified as a key efficiency category. Thus, temporal efficiency was isolated and studied in a laboratory environment using a single task being performed by participants.

2.2.4.3 Measuring user satisfaction

User satisfaction measures how well-designed the pilot thinks the low-cost CAS and its HMI is for the task of locating other traffic. This subjective assessment can be quantified using the System Usability Scale (SUS) [23]. The SUS score is determined using a 10 question Likert scale, administered to the user after having worked with the system. Adjective interpretation guidance for SUS scores was provided by Bangor, Kortum, and Miller [12]. The SUS score can be decomposed into two subscores [96]. These are the usability score and the learnability score. These subscales respectively quantify the user's perceived effort in working with a system or learning how to use that system. The remainder of this thesis focuses on the SUS-derived usability and learnability scores as measures of user satisfaction.

2.3 Scientific novelties of the approach presented in this thesis

The thesis at hand attempts to contribute to the scientific body of knowledge in several ways. On a methodical level, it exchanges research methods between the fields of gliding research and human factors research (see Figure 2.7). Usability research methods, such as questionnaires and laboratory experiments, are applied to a technical system in gliding (chapters 4 and 5). While this is not new to aerospace research in general, gliding has hardly been studied using such human factors approaches. Also the mathematical methods of inferential statistics are uncommon there. The thesis at hand should be seen as a step beyond the efforts of Jarvis [90] and Morgenstern [102] at establishing human factors research methods in gliding.

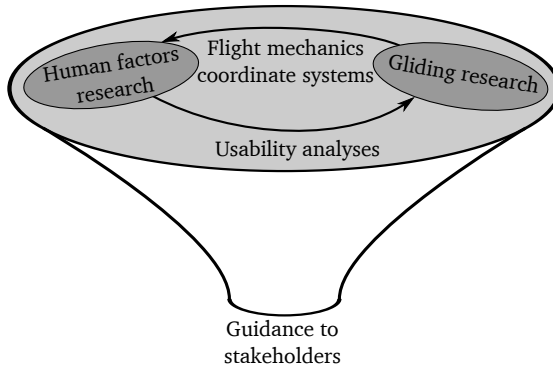


Figure 2.7.: Transfer of research methods in the thesis at hand. [illustration by author]

A transfer of research methods also exists in the opposite direction. In chapter 6, multiple coordinate systems are proposed which may form the basis of MMs developed by glider pilots. The concept of coordinate systems and coordinate transformations relies on the mathematical tools of linear algebra. These are often found in flight mechanics and gliding research, yet are rarely used in human factors work. This exchange of research methods is not an end in itself. Instead, these methods are pooled and funneled with the goal of answering the research questions of section 2.1.3.6. The answers gained using this approach allows for guiding recommendations about low-cost CASs to be extended to different stakeholders in the gliding community (chapter 7).

3 Human-machine interfaces of low-cost collision alerting systems in gliding

Low-cost collision alerting systems (CASs) must be equipped with some form of human-machine interface (HMI) in order to convey their information to the glider pilot. In the first part of this chapter, an overview on commercially available HMIs for low-cost CASs in gliding is provided. From this market overview it is evident that a perspective display format has not yet found its way into the glider cockpit. Therefore, a prototypical perspective traffic display format for gliding is proposed. This perspective display format also relies on a horizon-fixed camera attitude. It is conceptually designed in the second part of this chapter. At the end of this chapter the characteristics of this novel display format is compared to two typical commercial off-the-shelf (COTS) formats. This leads to multiple hypotheses about differences in usability.

3.1 Commercial-off-the-shelf human-machine interfaces

As already described in section 2.1.3, low-cost CASs are routinely found in gliders. Particularly the FLARM system forms a quasi-standard in the community. Due to the open data port architecture of the FLARM system [65], a multitude of different HMI devices exist. These devices rely on FLARM transceivers as traffic sensors. In the following section, a market study of COTS HMIs for the FLARM system is given. Selected HMIs relying on other sensing technologies or HMIs from prototypical research applications are also presented. These HMIs are categorized according to a newly proposed taxonomy, which is shown in Figure 3.1.

The taxonomy hierarchically categorizes HMIs of low-cost CASs according to several characteristics. On the highest level HMIs are distinguished according to the primary sensory channel used for relaying traffic information to the pilot.

Most products mainly rely on a visual depiction of traffic information while audio features are rarely used. However, in most cases, visual HMIs are supported by rudimentary audio alerts. This is in line with a recommendation of the European General Aviation Safety Team [57, p. 10]. At the second level, differentiation

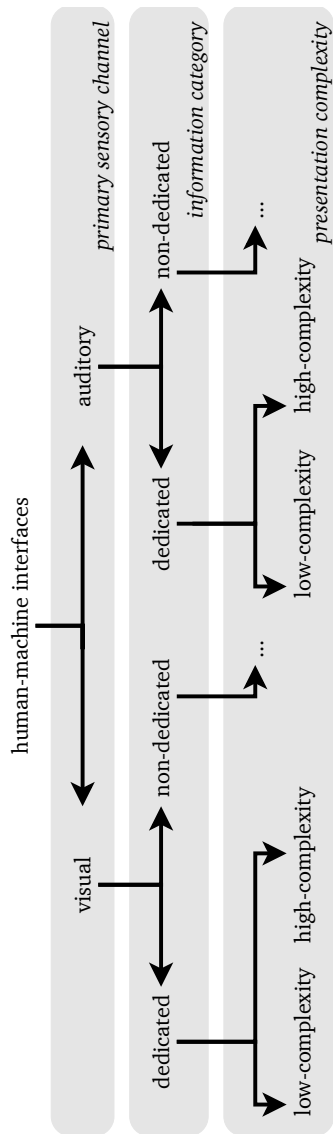


Figure 3.1.: Taxonomy for categorizing human-machine interfaces of low-cost collision alerting systems. [illustration by author]

is achieved according to the level of dedication. Some products are dedicated to the task of relaying traffic information to the pilot. Other products integrate this information along with other information, such as terrain and airspace warnings. Presentation complexity is used to differentiate HMI at the third level. Categorizing dedicated auditory and dedicated visual HMIs according to their presentation complexity leads to several clearly definable groups of COTS products. Non-dedicated HMIs are much harder to define using this approach, since the number of non-traffic-related data elements which could possibly be relayed to the glider pilot is quasi infinite. Therefore, these non-dedicated HMIs are only briefly discussed farther below.

Multiple definitions are introduced as the basis of the HMI taxonomy:

Definition 3: *Dedicated traffic HMIs* convey only information to the pilot regarding other aircraft in the vicinity of the pilot's ownship, as processed by a CAS, as well as information on the operating status¹ of the CAS.

Definition 4: *Non-dedicated traffic HMIs* integrate traffic information from a CAS along with information from arbitrary other systems.

Definition 5: *Low-complexity traffic displays* are dedicated traffic HMIs with a limited number of individually controlled lighted elements. The illumination of each individual element is linked to a unique meaning.² One additional information element - such as distance to the traffic - may be presented alphanumerically.

Definition 6: *High-complexity traffic displays* are dedicated traffic HMIs with a high number of individually controlled lighted elements. A meaning is not derived from the illumination of an individual element, but much rather from grouping illuminated elements to form symbols and alphanumeric characters.

Definition 7: *Low-complexity auditory traffic HMIs* are dedicated HMIs which relay only limited traffic information to the pilot via audio signals. This information is limited to the existence of other traffic and the severity of collision threats. It is relayed through rudimentary alert signals, such as gongs, chimes or beepers.

¹ The operating status of the CAS can be influenced by characteristics such as electrical power, information exchanged via the radio transceiver, or Global Positioning System position fix quality.

² For example, the illumination of a certain element may mean that traffic is located within a certain angular interval above the horizon.

Definition 8: *High-complexity auditory traffic HMIs* are dedicated HMIs which relay lingual traffic information to the pilot via audio signals. This information is not limited to the mere existence or threat level of other aircraft, but contains further information, such as the traffic's position or direction of motion. It is typically relayed via voice output.

Table 3.1 depicts those products studied during the market study. They are discussed more comprehensively in the following sections.

3.1.1 Visual human-machine interfaces

The vast majority of HMIs surveyed during the market study relies on the visual channel for relaying traffic information to the pilot. When these displays only show information on one or more aircraft in the vicinity of the ownship or on the operating state of the CAS they are classified as dedicated traffic displays. Non-dedicated traffic displays integrate different data types - including traffic information - from different sources [34, pp. 447–455].

3.1.1.1 Dedicated traffic displays

Those displays which are solely dedicated to the task of showing traffic information and status information of the CAS can be grouped into one of two categories. These categories are low-complexity displays and high-complexity displays.

Low-complexity displays

Low-complexity displays are display formats where a limited number of individually controlled lighted elements, such as light-emitting diodes (LEDs), is used. Each lighted element has a characteristic and unique meaning associated with its illumination. Commonly one- or two-dimensional position information about nearby traffic is provided in a polar coordinate system (CS). Only few of these displays provide full three-dimensional traffic positions including distance information.

Multiple low-complexity displays are commercially available. Every FLARM transceiver unit is equipped with a low-complexity display by default [64], as shown in Figure 2.2. ABOBA Elektronik's external FLARM Display V3 (Figure 3.2) is very similar in design to the FLARM transceiver's display [1]. The external display is used to replicate the FLARM transceiver's HMI, for example when the transceiver is remotely mounted outside of the pilot's field of view (FOV). For both displays,

Table 3.1.: Products reviewed during the market study.

Dedicated low-complexity displays	
FLARM Technology	FLARM transceiver unit (hardware version 3)
ABOBA Elektronik	external FLARM display V3
FLARM Technology	FLARM transceiver unit (legacy hardware)
EDIATec	ECW 100
Porod	patented collision alerting system relying on radio direction finding
Andrews	pilot warning instrument
Dedicated high-complexity displays	
Butterfly Avionics	external Butterfly display
LX navigation	Flarm Color Display II
Butterfly Avionics	PowerFLARM
Garrecht Avionik	TRX-2000
Zaon Flight Systems	PCAS MRX
Zaon Flight Systems	PCAS XRX
Non-dedicated displays	
LXNAV	LX9000
triadis engineering	Altair
Naviter	SeeYou Mobile
XCSoar project	XCSoar
Dedicated auditory human-machine interfaces	
triadis engineering	TR-DVS
Non-dedicated auditory human-machine interfaces	
triadis engineering	Vega
triadis engineering	FLOICE

information on the transceiver's operating mode, power supply and Global Positioning System receiver are shown on a LED bar on the displays' left-hand side. Also, position information regarding the traffic is shown in polar coordinates in both cases. A circular array of 12 LEDs is located at center of each display. It provides angular information on the traffic's bearing ρ , relative to the ownship's ground track. Angular information on the traffic's elevation ε relative to the ownship's altitude is given via a second bar of four LEDs. While this elevation bar is located at the right hand side of ABOBA Elektronik's external FLARM Display V3, it is at the center of the circular LED array of the FLARM transceiver box's integrated display of hardware version 3.

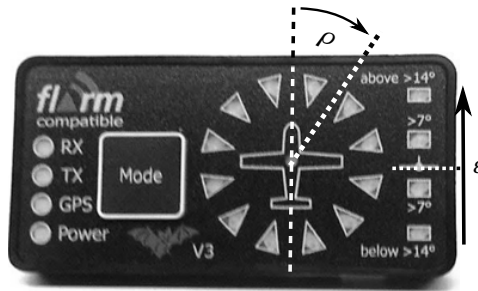


Figure 3.2.: Exemplary low-complexity display: The external FLARM display V3. [photograph and illustration by author]

Both displays show only one aircraft at a time. Depending on the operating mode, this is either the traffic nearest to the ownship's position or the traffic which is considered to be the highest collision threat. Whenever traffic is considered to be a threat, a collision alert is issued. This visual alert consists of flashing bearing and elevation LEDs. Their flashing frequency varies with the estimated time to collision. Furthermore, the visual alert is supported by a low-complexity audio alert drawing the pilot's attention to the display.

The orientation of the polar CS shown on these displays may not be as intuitive as suggested by the displays' simple design. According to FLARM's operating manual the relative bearing ρ is the angle between the ownship's ground track and the vector from the ownship to the traffic, when projected onto the horizontal plane [64]. The traffic's elevation ε is the angle formed by the vector from ownship to the traffic as it intersects the horizontal plane passing through the ownship.

An analogy from the field of navigation can be used to visualize the geometric interpretation of the relative bearing ρ and the elevation ε . In this analogy, the ownship is located at the center of a horizon- and track-fixed polar CS. The equa-

torial plane of the CS passes through the ownship's position and is parallel to the horizontal plane. Its zero-meridian passes through the direction of the ownship's ground track. The traffic's relative bearing ρ is analogous to longitude whereas its elevation ε is interpreted as latitude. Figure 3.3 visualizes this analogy. For a more precise mathematical description of these CSs and angular relations see appendix A in general and section A.2.1 specifically.

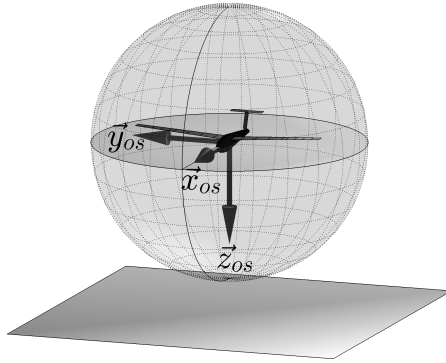


Figure 3.3: Visualization of the navigation analogy (ownship-fixed Cartesian and polar coordinate systems). [illustration by author]

When comparing the horizon- and track-fixed CS described within FLARM's operating manual [64] to the design features on the FLARM transceiver's integrated display (Figure 2.2) or ABOBA Elektronik's external FLARM Display V3 (Figure 3.2), a conflict of information can be identified. Both displays lack in-cockpit clues on proper orientation of the polar CS used to indicate the traffic's position. The CS orientation is only described within the respective manuals. Applying Roscoe's concept of pictorial realism [119] to the available in-cockpit clues of the low-complexity displays leads to a different polar CS which is suggested to pilots. Both displays are installed in a fixed position within the ownship. The fixed installation, without other in-cockpit clues, most likely suggests an ownship-fixed polar CS (Figure 3.4b). This ownship-fixed CS has an equatorial plane which is parallel to the ownship's wing plane and its zero-meridian passes through the ownship's longitudinal axis.

There are several design features directly on the two COTS low-complexity displays presented so far which further suggest an ownship-fixed CS. The FLARM transceiver's integrated display (Figure 2.2) lacks clues as to where the top of the circular LED array may be aligned. Due to its ownship-fixed installation, most often on the instrument panel and facing the airplane's longitudinal axis, it suggests that

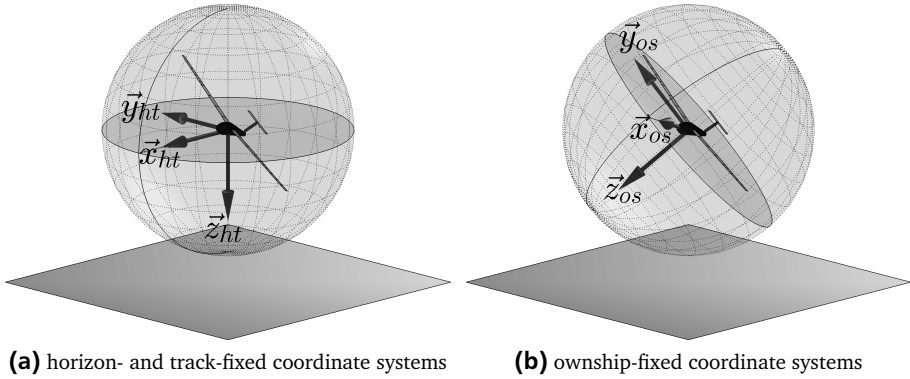


Figure 3.4.: Orientation of Cartesian and polar coordinate systems while maneuvering. [illustrations by author]

the top of the circular LED array is aligned with the longitudinal axis. This effect is even amplified in case of ABOBA Elektronik's external FLARM Display V3 (Figure 3.2) where the longitudinal axis of the ownship symbol, sketched at the middle of the circular LED array, points directly in the top direction of the LED circle.

For both displays, the orientation of the polar CS's equatorial plane is also incompletely or falsely explained by the in-cockpit clues. In case of the FLARM transceiver's display, no clues are given as to where the equatorial plane is located. Again, seeing the ownship-fixed installation, it seems prudent to assume that the equatorial plane of the polar CS would be parallel to the ownship's wing plane, when lacking knowledge from the FLARM manual. This orientation of the suggested reference plane is made even more evident in case of ABOBA Elektronik's external FLARM Display V3. Here, the ownship symbols at the center of the circular LED array and at the center of the right-hand vertical LED bar are sketched from a perspective directly normal to the wing plane or directly parallel to the longitudinal axis.

Particularly while maneuvering - and thus changing the glider's attitude away from straight and level flight - the suggested ownship-fixed polar CS (Figure 3.4b) does not coincide with the actual horizon- and track-fixed polar CS (Figure 3.4a). This is evident when comparing the sub-figures of Figure 3.4. Whenever the suggested ownship-fixed and actual horizon- and heading-fixed CSs do not coincide, it is suspected that pilots will make systematic errors in predicting the traffic's current position based on the information displayed.

Work performed by Aretz [8] and Aretz and Wickens [9] indicates that pilots mentally rotate displays. This mental rotation takes the display from the ownship-fixed CS and rotates it into the direction of the CS suggested by the display format. This mental rotation process demands time from the pilot and the temporal demand increases as the magnitude of the rotation angle increases [129]. In the case of the low-complexity display format, where an ownship-fixed CS is suggested, the temporal costs are not expected to vary between different flight conditions. This suspected indifference is due to the fact that the suggested CS of the low-complexity display format always remains aligned with the ownship-fixed CS.

Some legacy FLARM transceiver boxes feature an integrated low-complexity display lacking any elevation information on the traffic [64]. Therefore, they provide only one-dimensional information on the relative bearing to traffic. Contrary, EDIA-Tec's ECW100 is one of several low-complexity displays for FLARM which enhance the two-dimensional relative bearing and elevation information with an alphanumeric indication of the horizontal distance to the traffic [45].

Low-complexity display formats for collision alerting applications have already been proposed in the pre-FLARM era. Porod's CAS for gliders, described in a German patent, uses a circular LED array to show one-dimensional bearing information to the traffic [116]. Andrews analyzed a *pilot warning instrument* for collision alerting in powered aviation applications [5]. The instrument showed not only two-dimensional bearing and elevation information but also resolution advisories in a low-complexity format.

During a survey conducted within the experimental phase of this thesis 135 of 137 participating glider pilots indicated that they have flight experience with low-cost CASs as well as low-complexity display formats. Due to their widespread use, a COTS low-complexity display is used as baseline for comparing other display formats in the experiment of chapter 4. ABOBA Elektronik's external FLARM Display V3 [1] is selected as this baseline display. The suggested CS is presented more clearly on ABOBA Elektronik's display than on the FLARM transceiver's display. Also, the elevation signal should be easier to read on the ABOBA Elektronik display, as it is located outside of the circular LED array.

High-complexity displays

High-complexity displays are display formats which have a high number of individually controllable visual elements. These elements are arranged in monochromatic or polychromatic screens. The illumination and color of any individual pixel alone is not linked to a unique meaning. Instead, the meaning is derived from symbolic or alphanumeric representations drawn onto the screen using these pixels. Commercially available high-complexity displays are further categorized as

radar-style and alphanumeric displays. Both formats are discussed in more detail below.

Radar-style displays: Commercially available displays in the high-complexity category for gliding CASs exclusively rely on a camera position located far above the ownship to depict the traffic situation. A map of the traffic's relative horizontal position is drawn, similar to a radar screen used by air traffic controllers. Full three-dimensional data is provided for one or more traffic aircraft using symbols and text.

Two examples of commercial implementations of radar-style display formats for the FLARM system are the external Butterfly display (Figure 3.5) produced by Butterfly Avionics [25] and LX navigation's Flarm Color Display II [99]. Both products are similar in terms of design and suggested CSs. Due to the more comprehensive documentation available, the external Butterfly display is discussed as an example case in more detail.

In case of no traffic being categorized as a threat, traffic information is given in the radar-style display format. In this format, a dot forms the ownship symbol at the center of a two-dimensional polar CS. Two scale rings are located concentrically around the ownship symbol. The radius of the outer ring is shown alphanumerically at the screen's lower left-hand corner. The inner ring's radius is one half of the outer radius. The traffic symbol's angular coordinate in this polar CS represents its bearing ρ relative to the ownship's ground track whereas its radial coordinate represents the traffic's horizontal distance R_H . Alphanumeric information on one selected traffic aircraft is given on the screen's right-hand side. Whenever only one aircraft is detected by FLARM, that aircraft becomes automatically selected and its alphanumeric information is automatically displayed. This information includes the traffic's altitude ΔH relative to the ownship's altitude, the traffic's rate of climb \dot{H} , its abbreviated identification number and an alphanumeric repetition of its horizontal distance R_H . In case that a high threat traffic alert is issued, both displays – Butterfly Avionics' external Butterfly display and LX navigation's Flarm Color Display II – revert to emulating a low-complexity display format. These alerts are also enhanced with low-complexity audio signals, intended to draw the pilot's visual attention to the display.

The radar-style display format's external camera position and radar-style map implicitly suggest that it is invariant to changes in the ownship's pitch angle Θ and bank angle Φ . This implies a horizon-fixed equatorial plane of the polar CS. However, as is the case with the low-complexity FLARM displays, clues regarding the proper top direction of the polar CS are lacking. The radar-style display format therefore suggests a horizon- and heading-fixed CS, while they actually depict a horizon- and track-fixed system.

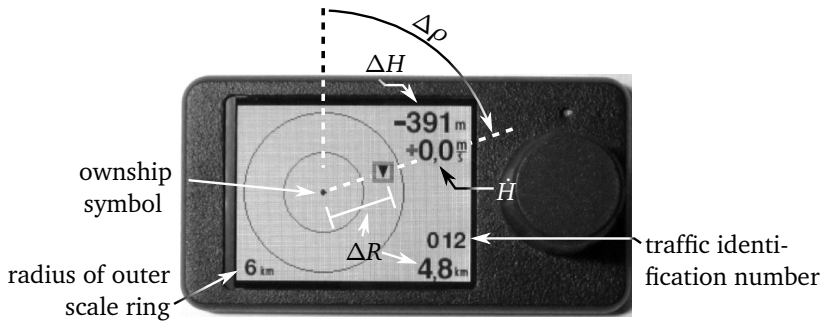


Figure 3.5.: Exemplary radar-style display: The external Butterfly display. [photo-graph and illustration by author]

The radar-style display format's suggested CS (horizon- and heading-fixed) and actual CS (horizon- and track-fixed) always coincide except when the airplane experiences a difference between heading and track. This difference is known as the drift angle ν and exists in steady-state flight conditions when a crosswind acts on the ownship or when a sideslip is being performed. Under these conditions it is assumed that pilots using the radar-style high-complexity display will perform systematic errors when searching for traffic.

In gliding, the radar-style display format is the second most-popular format of dedicated traffic displays. However, radar-style displays are much less popular than low-complexity displays. Only 42 of 137 glider pilots surveyed responded that they have flight experience with radar-style display formats. Nevertheless, this format is considered to be important for comparison. Particularly cross country and contest pilots use this format due to tactical advantages gained with seeing the rate of climb of other gliders nearby. Due to this popularity among highly experienced glider pilots, along with the fact that the radar-style display format suggests a CS already having a proper equatorial plane, it is decided to include a radar-style display in the experimental evaluation of this thesis. For reasons of easy availability and ease of installation, Butterfly Avionics' external Butterfly display [25] is selected as the second COTS display for experimental evaluation. Its characteristics are representative of the class of radar-style high-complexity displays commercially available for the FLARM CAS.

Contrary to the low-complexity display format, the CS suggested by the radar-style display format is generally not aligned with the ownship-fixed CS. It is suspected that when using the radar-style display format, mental rotations are performed by glider pilots. These rotations should account for deviations in pitch

angle Θ and bank angle Φ away from straight and level flight. This implies that pilot response times are suspected to increase whenever non-zero pitch or bank angles exist. At the same time, response times should be indifferent to changes in drift angle ν .

Also, it is assumed that response times are longer when using the radar-style display format than when using the low-complexity format. On the low-complexity format, the traffic's elevation ε is directly presented in form of an illuminated LED. The radar-style display format lacks such explicit angular elevation information. Instead, the elevation must be mentally estimated by the pilot using the alphanumeric depiction of relative altitude ΔH and through the symbolic or alphanumeric indication of horizontal distance R_H to the traffic.

$$\tan \varepsilon = \frac{\Delta H}{R_H} \quad (3.1)$$

Performing this trigonometric mental calculation is assumed to be time-consuming. Also, response speeds are quicker when data is presented graphically compared to textual representations [147].

Radar-style display formats are also found in products for powered aviation. For example, Butterfly Avionics' PowerFLARM portable CAS or Garrecht Avionik's TRX-2000 CAS rely on this format [26; 70]. However, characteristics such as symbols, units of measure and parameters shown are different for the use case of powered aviation.

Alphanumeric displays: Alphanumeric displays mainly use letters and digits to convey information about nearby traffic. Also, a limited number of rudimentary symbols, such as arrows, may be used. This display format is usually implemented on a monochromatic matrix screen. A schematic representation of an alphanumeric display format is given in Figure 3.6.

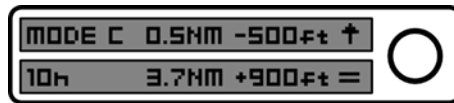


Figure 3.6.: Schematic depiction of an alphanumeric display. [illustration by author]

The alphanumeric display format enjoys limited popularity for powered-aviation collision alerting purposes. Zacon Flight Systems, now out of business, produced the PCAS MRX and PCAS XRX collision alerting units [161; 162]. Both units have an alphanumeric display and have gained some popularity within the powered aviation community in the United States. However, these systems do not use Global Posi-

tioning System position broadcasting as means of sensing traffic. Instead, they rely on transponder signal analysis and radio direction finding. Individual reports of glider pilots using these systems under special circumstances have been obtained. Apparently, these systems are used to separate gliders from powered aircraft operating in the vicinity whenever FLARM is not in widespread use.

Within the German gliding community this display format is not widely used for collision alerting purposes. Only 2 of 137 participating glider pilots have flight experience with this format. Due to this low-spread proliferation, along with the disadvantage of long expected reaction times [147], the alphanumeric display format is not analyzed experimentally.

Perspective displays: None of the commercially available displays for CASs in gliding yet rely on a camera position located at the pilot's point of view. They do not form a perspective representation of the out-the-window view, as the pilot would see it. The thesis at hand attempts to close this feature gap by designing a perspective display and evaluating its potential for gliding applications (section 3.2).

3.1.1.2 Non-dedicated displays

Non-dedicated traffic displays integrate traffic data with other data of interest to the glider pilot. Any HMI not only conveying information on the status of the CAS and on traffic, but also showing arbitrary other information, is considered non-dedicated. The format on how traffic is depicted in non-dedicated displays varies with the nature of other data displayed. This other data can range from navigational information on topography, airfields and airspace, to tactical tools such as final glide calculations and contest task surveillance. Due to this heterogeneity, extending the taxonomy to sub-levels of non-dedicated displays is inadequate. Because the focus of this thesis is on the interpretation of traffic representations, and not on data integration, non-dedicated HMIs are excluded from the experimental study. They are only discussed briefly in this market study.

During the market study two different categories of hardware implementations of non-dedicated traffic displays were found. These were either permanently installed glide computers or portable hardware solutions. Permanently installed glide computers are often integrated into the instrument panels of high-performance gliders. Some of these computer systems, such as LXNAV's LX9000 [98] or triadis engineering's Altair [143], can be interfaced with FLARM transceivers. The popularity of using integrated glide computers as non-dedicated traffic displays is similar to the use of radar-style high complexity displays for the task. 42 of 137 participating glider pilots reported that they have flight experience with such units.

Whenever the effort of permanently installing an integrated glide computer is too high, pilots can revert to portable solutions. In these cases, personal data assistant computers or mobile telephones running specific glide computer software are typically used. Payware and freeware applications - such as Naviter's SeeYou Mobile [106] or the application from the XCSOAR project [160] - can also depict FLARM-sensed traffic on a multitude of different hardware platforms. The use of portable hardware solutions displaying non-dedicated traffic displays is more popular than using integrated glide computers for this purpose. 52 of 137 queried glider pilots have flight experience with such portable solutions.

3.1.2 Auditory human-machine interfaces

Many visual HMIs for FLARM are enhanced with audio signals of varying complexity. However, not many COTS HMIs for FLARM rely primarily on the audio channel. The market study revealed only one manufacturer - triadis engineering - offering dedicated and non-dedicated auditory HMIs for gliding applications. TR-DVS is a product dedicated solely to providing high-complexity auditory traffic information to the glider pilot [146]. The second product, Vega, provides similar capability but also acts as an audio variometer, thus being a non-dedicated auditory HMI [145]. For use in powered aircraft the manufacturer also offers a third product, FLOICE, which is in use in several rescue helicopters [144]. All three systems convey high-complexity traffic information as voice outputs to the pilot. Whenever a higher-threat traffic alert is issued, this information is enhanced with a low-complexity audio alert.

triadis engineering is the only manufacturer of dedicated and non-dedicated auditory HMIs for FLARM. A query of German aircraft supply vendors reveals that the manufacturer's products can only be procured through the manufacturer directly. Because they are not sold through the regular sales channels and are built-to-order, these HMIs were considered to be uncommon. Glider pilots were not queried about the popularity of auditory HMIs and auditory HMIs were not analyzed experimentally.

3.2 Closing the market gap: Developing a perspective traffic display in gliding

The preceding market study revealed that all dedicated FLARM traffic displays commercially available are either low-complexity formats or use a radar-style camera position that is far above the ownship. None of these displays use a camera position which is located at the pilot's point of view, known as an egocentric camera

position. Such a position results in a perspective representation of features located in the environment. According to a literature review performed by Wickens and Prevett, egocentric display formats allow for superior pilot performance for local guidance tasks, such as visually identifying and avoiding nearby traffic [156].

Egocentric display formats for example can be found in commercial and military aviation applications such as synthetic or enhanced vision systems [34, pp. 462–465; 117]. In these applications the display's virtual camera is not only located at the pilot's eye point, but is also virtually fixed to the airframe. These truly egocentric displays provide either a virtual or enhanced out-the-window view to the pilot. They usually have a limited FOV to minimize geometric distortions which might hamper interpretability.

Due to the literature review of Wickens and Prevett [156], it is suspected that glider pilots would benefit from a CASs display format with increased egocentricity. It is the goal of the dissertation at hand to prove that usability will actually increase when glider pilots are presented with traffic information in a perspective manner. Therefore, such a display format is devised and implemented prototypically for further investigation. This novel display format should at least have increased egocentricity compared to the COTS display solutions available, provide a larger field of view than traditional synthetic vision systems and require only parameters from the FLARM serial data stream [65].

Möller, Kostka, Neujahr, and Klingauf also require increased egocentricity and a large FOV for their panoramic environmental display for a military fighter airplane [101]. Due to the distorting effects associated with fixing their projection's camera to the ownship, they gave up true egocentricity and reverted to a horizon- and heading-fixed camera attitude. This format was then called a heterocentric perspective display due to its non-ownship-fixed camera attitude. However, Möller et al.'s application had the advantage of being always providing the ownship's heading through the fighter airplane's attitude and heading reference system.

FLARM currently lacks such heading reference (compare section 2.1.3.2). Therefore the proposed perspective traffic display format for gliding applications must rely on a horizon- and track-fixed camera attitude, making it also heterocentric. Naturally, by using a track-fixed CS instead of heading-fixed CS on a perspective display format, concerns about increased pilot workload and decreased performance arise. However, track-fixing perspective depictions of the environment have not negatively impacted pilot performance and workload in a turboprop-powered general aviation aircraft [94]. The respective track-fixed synthetic vision system was even certified.

3.2.1 Conceptual design and prototypical implementation of a perspective traffic display

The traditional technology-centered approach to designing new systems and HMIs is prone to running into human performance limitations. This is partially due to users being exposed to the system or HMI during later stages of the design phase, where changes are cumbersome and expensive [48, chap. 1]. In aviation contexts, where operating near the human performance limitations can have catastrophic results, user-centered design processes are becoming more popular [91]. The key to user-centered design is including potential users in the design process from the beginning. Naturally, these users are initially exposed to non-perfect prototypes of a system at a lower technology readiness level (TRL) [56, Annex G] than their counterparts participating in technology-centered design. Yet it is this early exposure that helps the system designer shape the system and HMI around the user's subjective and objective needs for their task.

3.2.1.1 Design process

User-centered design is achieved by involving potential users throughout all stages of the design process. Literature, however, does not provide guidance for applying user-centered design to gliding use cases. Thus, the design process (depicted in Figure 3.7) leading to the prototypical perspective display was developed as part of the research at hand. According to an international standard, the steps undertaken in this process are only early design steps [132]. Over 1400 potential users were involved in the design and evaluation of the perspective display format's prototype, most of them in an online survey [80]. Thus, the corresponding design process exhibited a high degree of user-centricity. In comparison, the number of users involved in the technology-driven design processes of the two COTS display formats (low-complexity and radar-style) was about two orders of magnitude lower.³ Those design processes had a much lesser degree of user-centricity.

The design process leading to the prototypical perspective display consisted of four design phases (Figure 3.7). All phases, except the initial background research phase, included potential users in the design process.

Before beginning the display design, a design goal was defined: The design process should lead to a prototypical perspective traffic display for gliding which can be used in a simulator study comparing the new display format to selected COTS HMIs. Before commencing early experimental analysis, it should have TRL 3

³ M. Förderer, personal communication, September 24, 2012

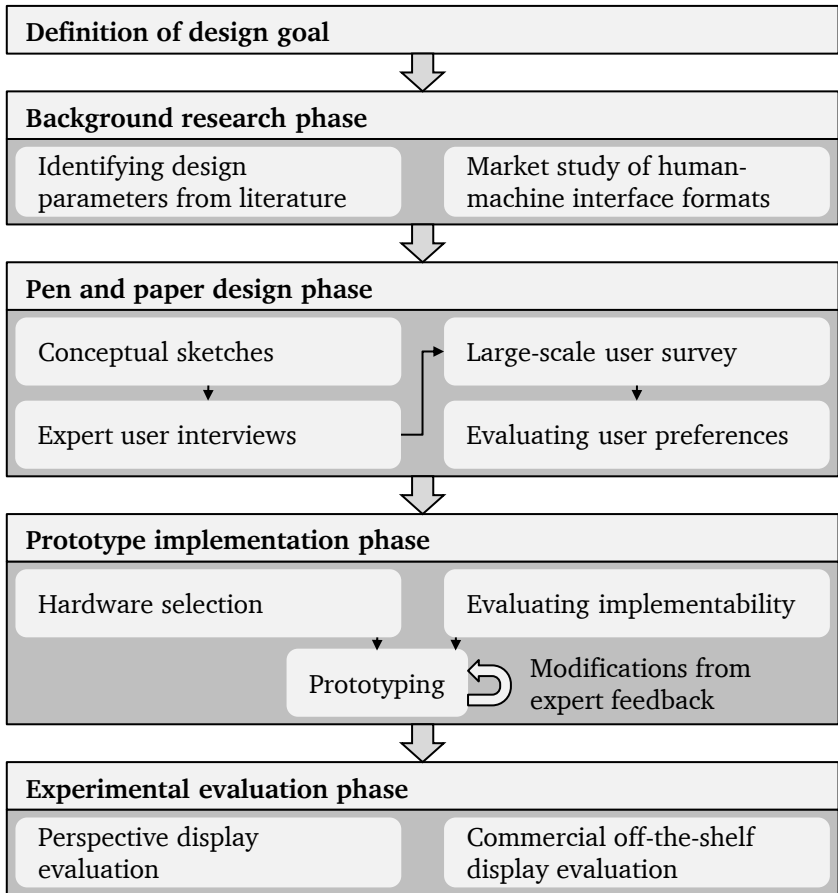


Figure 3.7.: Structure of the perspective display format’s design process. [illustration by author]

(“experimental proof of concept” [56, Annex G]). If the new format’s suspected benefits are supported by objective experimental results its TRL will increase to TRL 4 (“technology validated in lab”).

Background research phase

After defining the design goal, the process started with an initial background research phase. The market study on COTS HMI was part of this phase and was

presented in the preceding section 3.1. Schochlow and Gerber contributed to this phase with an unpublished review on projection formats and design parameters [126].

Pen and paper design phase

Heinbücher contributed to the following pen and paper design phase [80]. Taking the results from the preceding background research as well as the existing hypotheses, she developed multiple conceptual sketches of perspective displays. Three subject matter expert pilots, including two pilots of Technische Universität Darmstadt's research flight department provided her with feedback during expert interviews. Their feedback was used to modify the sketches, which were then presented to a wide range of German glider pilots by means of an online survey. A total of 1380 glider pilots provided their opinions on the conceptual displays via the online survey. Despite the fact that user-centered design is not "asking users what they want and then giving it to them" [48, p. 6], the large-scale user survey was a means of finding display presentations which are accepted in the community. By focusing on accepted display formats, ideally pilot performance and satisfaction will increase. Heinbücher then analyzed pilot preferences and recommended several display formats for implementation.

Prototype implementation phase

These recommendations formed the basis of the prototype implementation phase. At the beginning of this phase, each recommendation was evaluated whether it was implementable or not. Necessary modifications were also made. A prototype format was then implemented on adequate hardware for the experimental evaluation in a flight simulator. While implementing the prototype, three subject matter experts were exposed to the prototypes in expert interviews. Two of the experts were again pilots from Technische Universität Darmstadt's research flight department and the third expert was a cognitive psychologist with a background in usability research. Their suggestions were incorporated into the design of the prototype.

Experimental evaluation phase

Afterwards, the experimental evaluation phase followed. Here, 48 glider pilots were exposed to the prototype in a laboratory experiment. It was designed and performed together with Schochlow [125] and Mehringskötter [100]. The experiment's details are described in chapter 4. In order to compare the prototype to contemporary COTS display formats, an additional 89 glider pilots were exposed to one of two COTS HMIs in the same experiment. Comparing the experimental

usability results from the three displays (chapter 5) allowed for the prototype's TRL to be evaluated.

3.2.1.2 Display characteristics

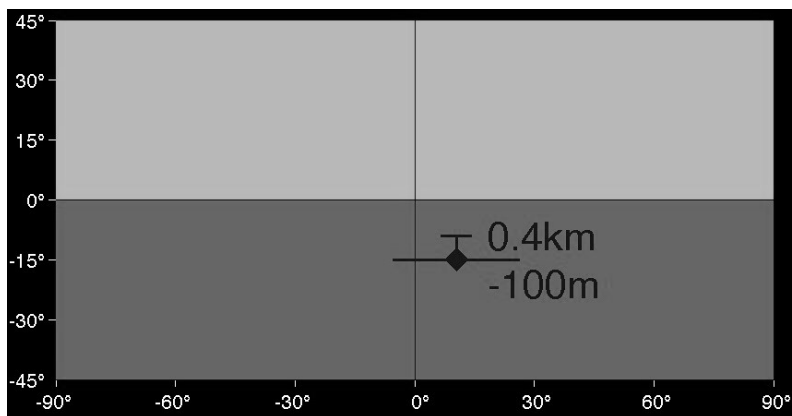
The prototypical perspective display format is illustrated in Figure 3.8. It was displayed on an LG Electronics D605 Optimus L9II mobile phone with a resolution of 853×480 pixels [97]. The mobile phone had a diagonal screen size of 11.9 cm.

Traffic was displayed by means of a symmetric cylindrical projection with a horizontal FOV of 180° and a vertical FOV of 90° . The horizontal axis represented the bearing ρ relative to the ownship's ground track while the elevation ε was represented along the vertical axis. Both axes were scaled equally with $19.0^\circ/\text{cm}$ and were marked with ticks and text labels. A marker cross indicated the zero values for ρ and ε . The display's top half had a light blue background color for elevations above the horizon ($\varepsilon > 0$) whereas areas below the horizon ($\varepsilon < 0$) had a light brown background. The horizon line remained constant on the display screen, irrespective of changes in the ownship's attitude. It did not act as an artificial horizon. This fixed horizon line served to suggest a horizon-fixed CS for the perspective display format. Since no heading information is available from the FLARM data stream [65], no heading clues were displayed and it was centered relative to the ownship's ground track. Therefore, the perspective display still had a disparity. Its clues implied a horizon- and heading-fixed CS while actually depicting a horizon- and track-fixed CS.

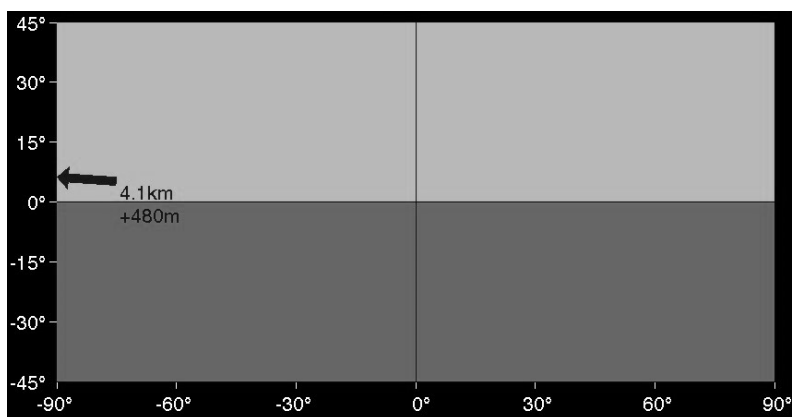
Traffic received by the system was displayed symbolically. All traffic located within the display's FOV (Figure 3.8a) was marked by a stylized glider symbol. Whenever the traffic was outside the display's FOV, an arrow symbol placed along the display's edge and pointed from the display's center to the traffic's hypothetical two-dimensional position. Text labels representing the traffic's horizontal distance R_H and relative altitude ΔH were placed in the direct vicinity of the glider and arrow symbols.

In order to improve the perspective impression, the sizes of the glider and arrow symbols as well as the associated text labels were scaled according to the traffic's slant distance R . Nearby traffic was displayed larger than traffic farther away. The scaling parameter k had the following form:

$$k(R) = \begin{cases} 1, & \text{if } R > 4000 \text{ m} \\ 2 + \frac{500 \text{ m} - R}{3500 \text{ m}}, & \text{if } 4000 \text{ m} \geq R \geq 500 \text{ m} \\ 2, & \text{if } R < 500 \text{ m} \end{cases} \quad (3.2)$$



(a) traffic in field of view



(b) traffic outside of field of view

Figure 3.8.: Perspective display format. [illustrations by author]

All dynamic data elements shown on the display format were derivable from the FLARM serial data stream [65].

Inherently, the prototype still had practical limitations which restricted its use to laboratory tests. The display code was implemented within the laboratory software program described in section 4.3. It was not a stand-alone software program for flight operations. Also, the prototype display only showed one traffic aircraft at a time using only glider symbols. Multiple traffic aircraft of different aircraft

categories or higher threat aircraft were not introduced. Audio features were not implemented. The hardware was selected for good readability in the simulator's low ambient light conditions. Problems associated with the high ambient light conditions found during a daytime visual flight still need to be addressed for practical flight operations.

3.3 Hypotheses

In the discussion of the market study (section 3.1), two concepts from literature were used to hypothesize pilot behavior. The first was the concept of pictorial realism [119]. It stated that each graphic display format implicitly suggests a CS which is used to display the data. Whenever the suggested CS and the display's actual CS do not coincide, errors are expected to arise. These errors can be quantified by showing traffic information on a display - but not in the out-the-window view - and then asking the pilot where the traffic is suspected. The angle $\Delta\gamma$ between the traffic's actual direction, as seen from the ownship, and the direction suspected by the pilot is a measure of error magnitude. Since the suggested CS varies with the display format, the following hypothesis was postulated:

Hypothesis 1: Visual search error magnitude $\Delta\gamma$ increases whenever the disparity between a display format's suggested and actual coordinate system exists. Larger disparities will lead to larger error magnitudes. Thus, display formats suggesting an ownship-fixed coordinate system (low-complexity display) will result in larger error magnitudes than display formats suggesting a horizon- and heading-fixed coordinate system (radar-style and perspective displays).

The second concept introduced during the market study was the concept of mental rotations [8; 9]. The corresponding train of thought implies that pilots must mentally transform data displayed to them from the CS suggested by the display format into their ownship-fixed seating position. The time required to perform these rotations increases with the magnitude of the rotation. Therefore, large disparities between the ownship-fixed and suggested CS will result in longer pilot reaction times t_R when interpreting the display than small disparities. This resulted in a further hypothesis:

Hypothesis 2: Reaction time t_R increases whenever a disparity between a display format's suggested coordinate system and the

ownership-fixed coordinate system exists. Larger disparities will lead to a longer reaction time. Display formats suggesting an ownership-fixed coordinate system (low-complexity display) will see a shorter reaction time than display formats suggesting a horizon- and heading-fixed coordinate system (radar-style and perspective displays).

Additionally to the objective parameters of visual search error magnitude $\Delta\gamma$ and reaction time t_R , subjective parameters may also be influenced by display design. HMIs developed by applying user-centered design will likely lead to higher subjective user satisfaction than COTS HMIs developed using more traditional design approaches. Subjective user satisfaction can be quantified using the System Usability Scale and its subscales (section 2.2.4.3). The usability subscore can be directly interpreted as the user's satisfaction. The learnability subscore provides a quantification of effort experienced by the user in learning to use a system. When the HMI designer has a good and thorough understanding of the user's level of knowledge, then learnability subscores are suspected to benefit. User-centered design is one method to provide the HMI designer with this required feedback on the user's level of knowledge. Because the design process of the perspective display format showed a much higher degree of user-centricity than the design processes of the two remaining COTS display formats, a third hypothesis was formulated:

Hypothesis 3: Subjective user satisfaction is higher for display formats whose design process has a high degree of user-centricity (perspective display). Display formats exhibiting a lower degree of user-centricity in their design processes (low-complexity and radar-style displays) result in lower subjective user satisfaction. This applies to the aspects of subjective usability and learnability.

All three aforementioned hypotheses discuss influences on the different dimensions of usability (see Figure 3.9). Pilot effectivity is subject of Hypothesis 1 while temporal efficiency is indirectly discussed when combining Hypotheses 1 and 2. The third usability dimension of user satisfaction is addressed in Hypothesis 3.

The hypotheses assumed that the displays and their associated characteristics, such as their suggested CSs, influence their respective usability. Yet these display characteristics were a result of the respective display design processes. Therefore, the three individual hypotheses were summarized in one global hypothesis:

Global Hypothesis: Increasing user-centricity in the display design process results in HMIs which provide optimized usability in all three usability dimensions.

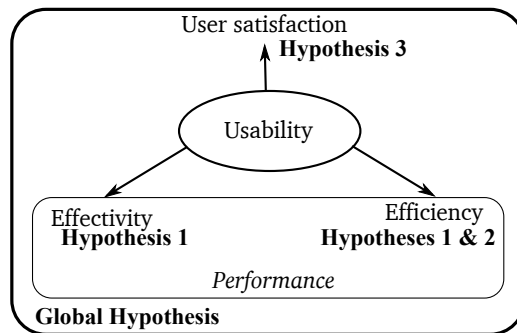


Figure 3.9.: Hypotheses and corresponding usability dimensions. [illustration by author]



4 Determining the influence of display formats on usability

In order to test the usability hypotheses of section 3.3, an experimental design is presented in the current chapter. The methodical design is discussed by introducing independent and dependent variables. Following this, the sample of participants is characterized. Afterwards, hardware and software used during the experiment are presented and the procedure and task performed by the participants are introduced. The chapter closes with a description on how the collected data is treated for further analysis.

At the onset of this study there was little a priori knowledge about how glider pilots interact with the human-machine interfaces of low-cost collision alerting systems (CASs). This made it difficult to anticipate the strength of the effects which were to be measured. In order to detect potentially small effects, it was decided to design an experiment in a highly controlled laboratory environment. Participants only performed a single task during the resulting experiment, which consisted of interacting with the display of a CAS without actually flying the simulator. Alexander and Wickens relied on a similar no-fly-task experiment in their evaluation of traffic displays by pilots of powered aircraft [2]. Naturally, such a highly controlled laboratory experiment reduces the transferability of the results to actual flight operations. However, it allows effects to be uncovered which may otherwise be masked by the varying environmental conditions of a less controlled and more realistic fly-task experiment.

4.1 Experiment design

The hypotheses of section 3.3 formed the basis for the experimental analysis and inherently shaped the experiment's design. Each hypothesis focused on a different usability dimension and its potential influences. Measurable parameters quantifying these different usability dimensions were selected as the experiment's dependent variables. Those factors which were hypothesized to influence the usability dimensions were defined as the corresponding independent variables.

Display format was an easily modified independent variable. Each participant was assigned to one of the three display formats for the experiment. The display

format was defined as an independent between-subjects variable with three factor levels. These factor levels were *low-complexity*, *radar-style*, and *perspective*.

Embedding the different relative orientations of the suggested and actual coordinate systems (CSs) in an experimental context was more cumbersome. This was achieved by exposing the participants to different relative orientations, which were narratively described as flight conditions. The flight condition was categorized as an independent within-subject variable with four factor levels. These four factor levels were *straight and level flight*, *horizontal flight with crosswind from the left*, *left-hand turning flight*, and *climbing flight*. Different factor levels of flight condition resulted in different disparities between FLARM's actual horizon- and track-fixed CS on the one hand and the suggested CSs of the different display formats on the other hand. *Straight and level flight* acted as a control factor level where all CSs coincided.

Each flight condition level differed in terms of bank angle Φ , pitch angle Θ and drift angle ν , as shown in Table 4.1. The relative orientations between the actual horizon- and track-fixed CS and the CS suggested by each display format are the basis for evaluating Hypothesis 1. They are described mathematically¹ in Table 4.2. Whenever CSs coincided, the respective coordinate transformation was described by the 3×3 identity matrix I_3 . Hypothesis 2 can be evaluated by looking at how the CS suggested by each display format is oriented relative to the ownship-fixed CS. How the orientation of these different CSs varies with changes in flight condition is described in Table 4.2. The factor levels of flight condition were selected so that the ownship-fixed and horizon- and track-fixed CSs were rotated relative to one another by different orthogonal axes and at different angular magnitudes (see Table 4.3).

The experimental task, described in the following section 4.5, was repeated 25 times per flight condition (repeated measures) by the participants with separate signals shown on the display. This resulted in a $4 \times 3 \times 25$ within-between repeated measures design of the experiment.

During the experiment, participants answered to where they expected the traffic shown on the display to be located in the out-the-window (OTW) view. Their assumed traffic locations and reaction times were recorded as the experiment's dependent variables. Since no traffic was shown to participants in the OTW view, their answers were solely based on the interpretation of the stimulus shown on the display. This prevented participants from gaining feedback on how they performed during the experiment. Lacking feedback, participants were unable to adapt their behavior [133, pp. 71–73].

¹ see section A.2 for derivation of the respective coordinate transformations



Table 4.1.: Characteristics of different factor levels of flight condition. [illustrations by author]

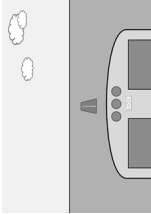
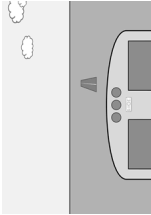
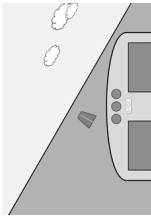
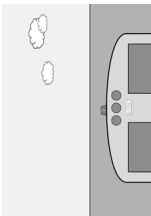
Flight condition factor level	<i>straight and level flight</i>	<i>horizontal flight with crosswind from left</i>	<i>left-hand turning flight</i>	<i>climbing flight</i>
Cockpit visual scene				
Bank angle Φ in $^{\circ}$	0	0	-30	0
Pitch angle Θ in $^{\circ}$	0	0	0	10
Drift angle ν in $^{\circ}$	0	-12	0	0

Table 4.2.: Relative orientation of coordinate systems during different levels of flight condition and display format (mathematical description).

Flight condition factor level	<i>straight and level flight</i>	<i>horizontal flight with crosswind from left</i>	<i>left-hand turning flight</i>	<i>climbing flight</i>
Low-complexity display format¹				
$M_{os,ht} =^3$	I_3	$\begin{bmatrix} \cos 12^\circ & -\sin 12^\circ & 0 \\ \sin 12^\circ & \cos 12^\circ & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos 30^\circ & -\sin 30^\circ \\ 0 & \sin 30^\circ & \cos 30^\circ \end{bmatrix}$	$\begin{bmatrix} \cos 10^\circ & 0 & -\sin 10^\circ \\ 0 & 1 & 0 \\ \sin 10^\circ & 0 & \cos 10^\circ \end{bmatrix}$
$M_{os,os} =^4$	I_3	I_3	I_3	I_3
Radar-style and perspective display formats²				
$M_{hh,ht} =^3$	I_3	$\begin{bmatrix} \cos 12^\circ & -\sin 12^\circ & 0 \\ \sin 12^\circ & \cos 12^\circ & 0 \\ 0 & 0 & 1 \end{bmatrix}$	I_3	I_3
$M_{hh,os} =^4$	I_3	I_3	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos 30^\circ & \sin 30^\circ \\ 0 & -\sin 30^\circ & \cos 30^\circ \end{bmatrix}$	$\begin{bmatrix} \cos 10^\circ & 0 & \sin 10^\circ \\ 0 & 1 & 0 \\ -\sin 10^\circ & 0 & \cos 10^\circ \end{bmatrix}$

¹ ownship-fixed CS suggested

² horizon- and heading-fixed CS suggested

³ transformation from actual horizon- and track-fixed CS used to depict data to the display format's suggested CS

⁴ transformation from ownship-fixed CS to display format's suggested CS



Table 4.3.: Relative orientation of coordinate systems (CSs) during different flight conditions (graphical description).
[illustrations by author]

Flight condition factor level	straight and level flight	horizontal flight with crosswind from left	left-hand turning flight	climbing flight
Orientation of ownship-fixed CS				
Orientation of horizon- and track-fixed CS				

4.2 Participants

In total, 137 glider pilots participated in the experiment. However, it was necessary to exclude the datasets of 16 participants from further analysis for different reasons. One participant who interacted with the low-complexity display showed signs of alcoholic intoxication during the experiment. Thirteen datasets of participants using the radar-style display rendered useless by an initial software recording error. Also, recalibrating the touch screen monitor (see section 4.3) became necessary during the sessions of two participants interacting with the perspective display. All of these unreliable datasets were purged from further analysis.

Of the remaining participants 121, a total of 8 were female and 113 were male. Descriptive statistics (mean M , standard deviation SD , minimum \min and maximum \max of each measured dependent variable) of the participants' demography and flight experience are provided in Table 4.4.

Table 4.4.: Demographics and flight experience of participants.

Parameter	M	SD	\min	\max
Age in years	27.7	13.8	14	71
Total flight time in gliders in hr	312.3	581.1	0.5	3710.0
Glider flight time within last 6 months in hr	13.8	20.4	0.0	150.0
Total number of glider flights	500.2	675.6	2	4067

Their glider pilot qualification also varied. Thirty-five participants were student pilots during their pre-solo or post-solo stages of glider flight training. Sixty-seven participants held regular glider pilot licenses without instructor privileges and 19 participants held instructor ratings with restricted or full privileges.

Participants were recruited by leaflets placed on the campus of Technische Universität Darmstadt and several airfields in the Darmstadt vicinity. Additionally, the experiments were advertised in social media groups and many participants became aware of the campaign by word-of-mouth from former participants. In total, participants came from 56 different home airfields. To avoid that participants would prepare for the experiment, no mention of CASs was made in these advertisements. Instead, only a study about “procedures in sport aviation” was advertised. The only

prerequisite was that participants have some glider flying experience. Because no additional prerequisites were made, low-time student glider pilots and inactive glider pilots were also allowed to participate. Also, participants knew beforehand that they would be compensated for their effort with free flight time in the Institute of Flight Systems and Automatic Control’s DA 40-180 flight simulator [86] after the experiment had ended.

Depending on when the participants took part in the experiment, they were assigned one of the three display format factor levels. Participants taking part between January and February of 2013 were assigned the low-complexity display. Experiments with the radar-style display were performed between April and May of 2014 while the perspective display was tested between December of 2014 and January of 2015. Differences in demography (age and sex) and flight experience between the participant samples assigned to the three display format factor levels were evaluated (see appendix C). No statistically significant differences between the groups exposed to the different formats were found.

The participants’ familiarity with their assigned display format varied, which is shown in Table 4.5. Fisher’s exact test performed on the data summarized in Table 4.5 revealed that familiarity varies significantly ($\chi^2(2) = 110.66, p < .01, V = .87$) with large effect size between the display format factor levels assigned. This effect was to be expected, since the low-complexity format is highly popular whereas the perspective display has not been used in CASs for gliding applications before.

Table 4.5.: Participants’ familiarity with their assigned display format.

Display format assigned	participants	
	familiar	unfamiliar
Low-complexity	41	1
Radar-style	8	25
Perspective	0	46

4.3 Equipment

The experiment was performed in the Institute of Flight Systems and Automatic Control’s Diamond DA 40-180 flight simulator [86]. It represented a single-engine piston-engine powered aircraft with a Garmin G1000 integrated avionics suite. Even though this cockpit environment was very different compared to the glider cockpits that the participants are normally exposed to, the simulator was used to

provide visual stimuli for immersion into the task. This was achieved by projecting the scenery of the OTW view onto a cylindrical projection screen with a field of view of $200^{\circ} \times 35^{\circ}$. The scenery was generated using the Diamond Global Canvas image generation software [38].

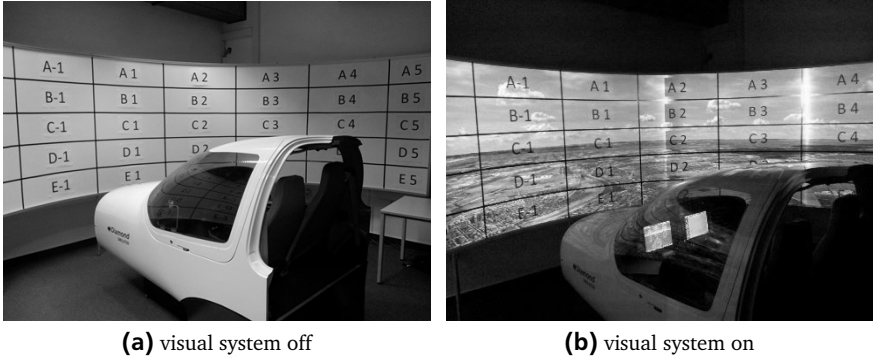


Figure 4.1.: Flight simulator cockpit mock-up with gridded projection screen. [photographs by author]

As seen in Figure 4.1, the 200° ownship-fixed horizontal field of view of the OTW view was separated into 10 columns. Similarly, the 35° ownship-fixed vertical field of view was divided into five rows. This created a grid of 50 cells in the OTW view, each of which was marked with a unique alphanumeric identifier. The OTW view grid allowed participants to easily announce where they suspected traffic to be located using a CTF840-SH touch screen monitor produced by CarTFT.com [28]. During the participants' experimental tasks, the monitor showed an answer grid mirroring the gridded OTW view (see Figure 4.2).

Depending on which display format was being evaluated by the participant, either the low-complexity display, the radar-style display or the perspective display was installed in the simulator. Each display was placed at the center of the simulator's instrument panel, between the primary flight and multifunction displays. The installed location of the displays is depicted in Figure 4.3.

Signals were generated and data was recorded on a Dell Precision M90 personal computer [37] running the Windows 7 Professional operating system. A self-developed FLARM Experiment Software recorded the participants' responses to each generated signal as well as the time between signal generation and the responses [62]. It interfaced with the touch screen monitor via video graphics array and universal serial bus connections.



Figure 4.2.: Touch screen monitor with answer grid. [photograph by author]

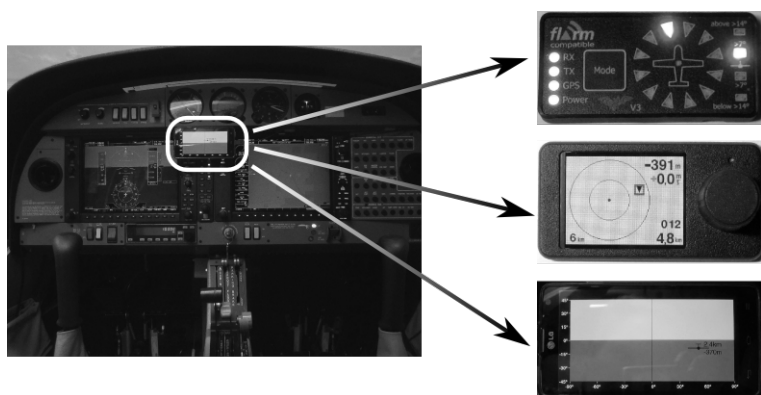


Figure 4.3.: Location of the installed displays in the flight simulator cockpit. [photographs and illustration by author]

How the actual display devices were interfaced with the personal computer varied between the types of displays (see Figure 4.4). For the commercial off-the-shelf low-complexity and radar-style displays, their own firmware was used to generate their graphics output. This output was based on a traffic signal received from the FLARM Experiment Software on a serial interface running the FLARM Technology serial protocol [62; 65].

The self-developed perspective display used a D605 Optimus L9II mobile phone as display hardware, running the Android 4.1.2 operating system [97]. Its graphics calculations were primarily performed by the FLARM Experiment Software on the personal computer [62]. Therefore, the mobile phone acted as an external monitor for the personal computer. It was connected to the personal computer via a universal serial bus connection. Both devices used the protocol provided by the iDisplay programs to communicate through this connection [85]. This allowed the mobile phone to be recognized as an additional monitor by the operating system of the personal computer.

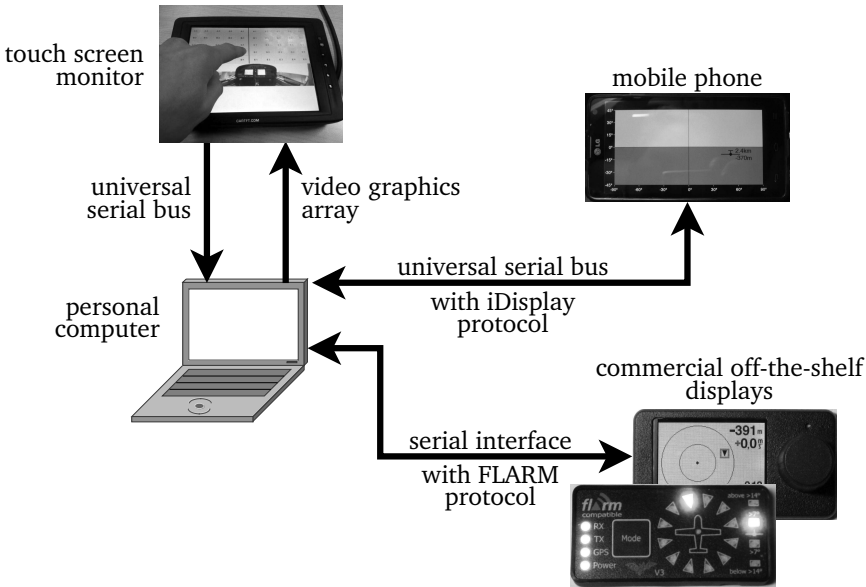


Figure 4.4.: Computer hardware and interface setup. [illustration and photographs by author]

Reaction times were measured using the FLARM Experiment Software [62]. Due to the software’s architecture, a software timer started as soon as a new signal was transmitted to the display device. The timer stopped as soon as the software recognized that the participant pressed a button on the touch screen monitor. Therefore, the software not only measured the time it took participants to react to a new stimulus on the display. It also included a component due to the hardware and software setup’s latency. In order to compensate for this latency, which occurred between the start of the software timer and the new signal being completely shown on the display, a video analysis was performed. The results, which are shown in Table 4.6, were then used to correct the recorded reaction times.

Table 4.6.: Results of video analysis to determine latency between start of software timer and new signal being fully shown on display device.

Display format	FLARM Experiment Software version	signals analyzed	<i>M</i> in ms	<i>SD</i> in ms
Low-complexity	1.22b	125	13.6	17.0
Radar-style	1.33	120	462.2	43.3
Perspective	1.40	120	32.5	18.5

The video analysis was performed with a recording frequency of 30 frames per second.

4.4 Procedure

While undergoing the experiment, participants took part in three consecutive sessions (Figure 4.5). These were the briefing session, the simulator session and the debriefing session. During the briefing session, which took place in a briefing room, participants were initially welcomed by an experimenter. They then participated in a general briefing on the nature of the experiment. In the briefing, participants also viewed pictures of the display format they would encounter in the simulator. At this point they were instructed that their display format visualized traffic information received by a FLARM transceiver. No further classroom instruction on the functions of the display and the CSs used was given. Instead, if unfamiliar with the display format, participants were asked to work intuitively. This lack of classroom instruction was expected to minimize interference with the participants’ mental models. During the briefing session participants also filled out two questionnaires, provided

information on their demography, flight experience and pilot qualifications as well as on their familiarity with CASs and their human-machine interfaces. At the end of the briefing session, participants signed an informed consent form. Underage participants also provided consent from a legal guardian before participating.

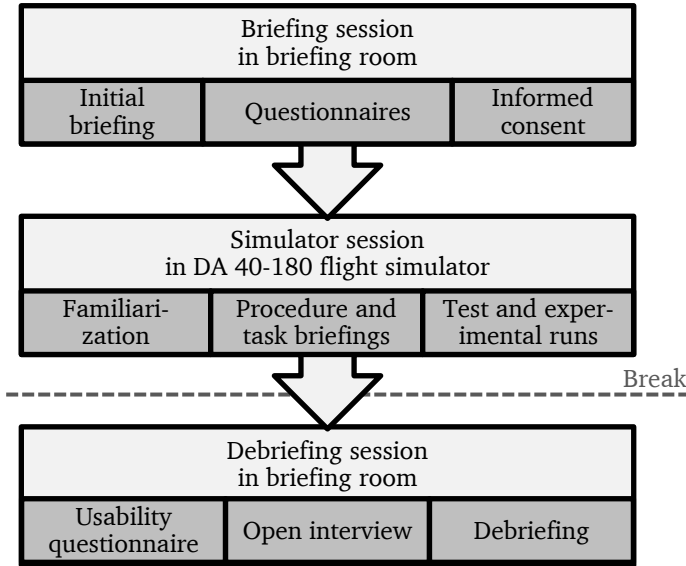


Figure 4.5.: Flow diagram of the experiment's sessions. [illustration by author]

Next came the simulator session. It was performed in the DA 40-180 flight simulator. Participants initially familiarized themselves with the simulator setup and the gridded OTW view. They then seated themselves on the left-hand seat of the simulator and placed the touch screen monitor on their lap. While seated, participants were briefed by the experimenter on the experimental procedure in the simulator and on the task that they were to perform.

The simulator was set onto the extended runway centerline of a hard surface runway at a distance of 3NM from the threshold and 2000 ft above field elevation. Then the experimental procedure, illustrated in Figure 4.6, commenced. Initially, the FLARM Experiment Software randomized the order of the flight condition's four factor levels [62]. The first level was then selected and the simulator's attitude was configured for the first factor level according to Table 4.1. Meanwhile, the simulator remained frozen. Participants then viewed a written description of the flight condition's factor level on the touch screen monitor. This description always

clarified that the simulator was tracking along the extended centerline of the hard surface runway visible in the OTW view. To compensate for the lack of optical flow, caused by the frozen simulator condition, participants then watched a video on the touch screen monitor. The video had a duration of approximately 15 s and showed a maneuver from the cockpit perspective of a light airplane flying at the given flight condition's factor level.

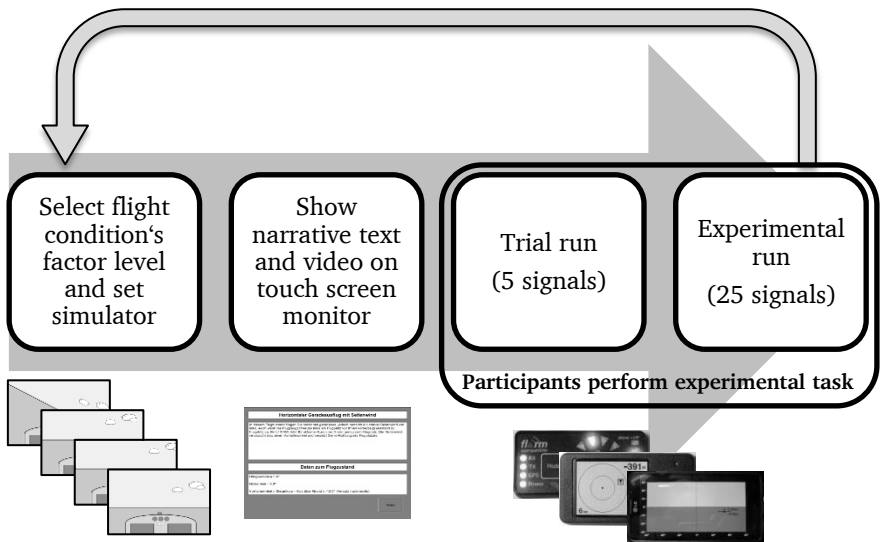


Figure 4.6.: Flow diagram of the experimental procedure in the flight simulator. [illustration and photographs by author]

In order to practice their experimental task, participants then performed a practice run. The FLARM Experiment Software selected 5 of the 50 cells of the gridded OTW view [62]. Invisible traffic aircraft were assumed to be located at the center of each of these five cells. For each invisible traffic aircraft, the slant distance was randomly sampled from a rectangular probability density function and varied between 200 m and 6000 m. Where necessary, the function's upper discontinuity was reduced to a value that bounded the traffic's position to within 500 m of the ownship's altitude. Display signals for each of these five aircraft were then calculated and shown at intervals of 9 s. Because signals were separated by a pause of 2 s when the displays were blanked, the traffic stimulus on the display was shown for

7s. During the practice run, participants' answers to their experimental task were not recorded. Before and after the practice run, participants could ask questions to clarify their task.

After these questions were clarified, the first experimental run commenced. During the experimental run, the FLARM Experiment Software selected 25 grid cells of the OTW view [62]. Correspondingly, the experimental run consisted of 25 signals being processed by the participants. Participants' answers were now recorded. Otherwise the practice and experimental runs were identical. After the first experimental run was finished, the experimental procedure was reiterated. The next factor level of flight condition was selected and the experimental procedure of Figure 4.6 was performed. This continued until the participants completed the experimental runs of all four factor levels.

After the simulator session participants took a break as long as they wished. Following this, the debriefing session commenced in the briefing room. There, participants assessed their display format using the System Usability Scale questionnaire. Before being debriefed on the purpose of the experiment, participants could express their opinions and experiences with their display format in an open interview. After the debriefing the experiment was finished. The overall duration of all three sessions was approximately 90 minutes. After the experiment was finished, most participants accepted their compensation and made use of free flight time in the simulator.

4.5 Experimental task

Participants performed their experimental task during the practice and experimental runs. The task is shown in Figure 4.7. It was briefed to participants at the beginning of the simulator session. The oral briefing consisted of the following instructions:

- a) Look at the display as soon as you notice a new signal. Interpret the signal.
- b) Glance to the OTW view where you suspect the indicated traffic to be located. Memorize the alphanumeric identifier of the OTW view's cell where you looked.
- c) Answer swiftly where your initial glance was directed by tapping the respective cell on the touch screen monitor.

Of these three sub-tasks, only the sub-tasks *a)* and *b)* are performed during actual glider flying. Sub-task *c)* was only inherent to the experiment. Schochlow proposed that the participants' experimental task could be described as a sequential

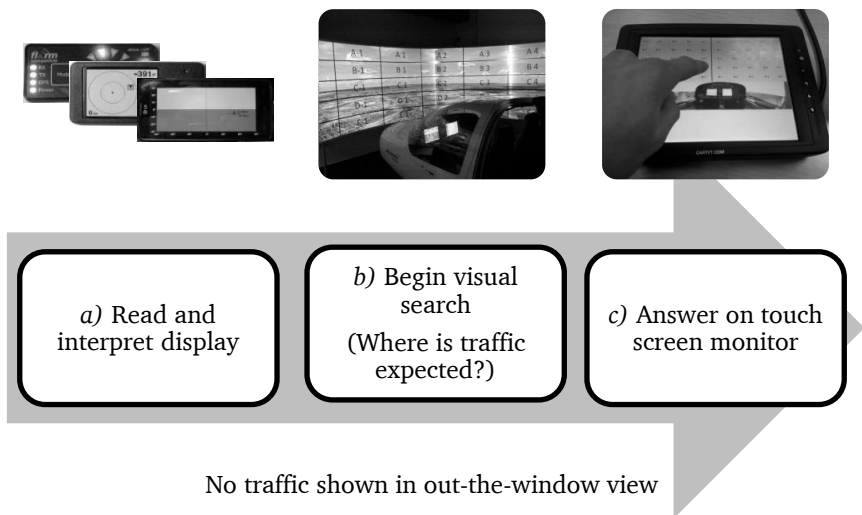


Figure 4.7.: Flow diagram of the participants' task during the trial and experimental runs. [illustration and photographs by author]

task model, consisting of the three sub-tasks *a)*, *b)* and *c)* [125, pp. 7–10]. According to the works of Donders [41] and Sternberg [135, 136] the time spent by each participant on the respective sub-tasks is cumulative. Therefore, when measuring the total time spent on all three sub-tasks, the time spent by participants on the flight-relevant sub-tasks *a)* and *b)* can be determined, if the time spent on sub-task *c)* is known. This is performed in detail in appendix D.

4.5.1 Feedback during task

Traffic was not shown in the OTW view. Showing traffic there was expected to skew participants' answers to where they *found* the traffic. However, the question of interest was where they *assumed* the traffic to be located.

The invisible traffic approach also implied that participants received no feedback whether their answers to the experimental task were correct. This lack of feedback precluded the participants from learning the CAS's actual behavior based on the signals shown on the displays [133, pp. 71–73]. Since the learning process was intentionally suppressed, the participants' mental models of the CAS were assumed to remain constant throughout the experiment.

4.6 Treatment of data and preliminary analysis

For each signal presented to the participants, the underlying position of the invisible traffic in the OTW view was also recorded. The position information consisted of the traffic's visual bearing τ and its visual elevation δ (see section A.2.2 for definitions of these angles). Several effectivity metrics were extracted from the discrepancies between the traffic's actual position and the participant's answer to where the traffic was assumed. One efficiency metric was the magnitude of the undirected angular error $\Delta\gamma$. It expresses how far away from the traffic's actual position the participants assumed the traffic to be located. Calculating the magnitude of $\Delta\gamma$ was achieved according to section A.3. The discrete nature of the gridded OTW view allowed these errors to be detected only at discrete levels.

Another efficiency metric was the participants' reaction time t_R for performing the flight-relevant experimental sub-tasks a) and b). It was extracted from the software-registered response time. Figure 4.8 illustrates that the software-registered response time consisted of two primary components: The first temporal component was the latency between the start of the software-timer during signal generation and the indication of the signal on the display (see section 4.3). The second component was the participant's time to complete the experimental task. Applying Schochlow's sequential task model [125], the software-registered response time was decomposed further into the times it took a participant to complete each of the three sub-tasks a) through c).

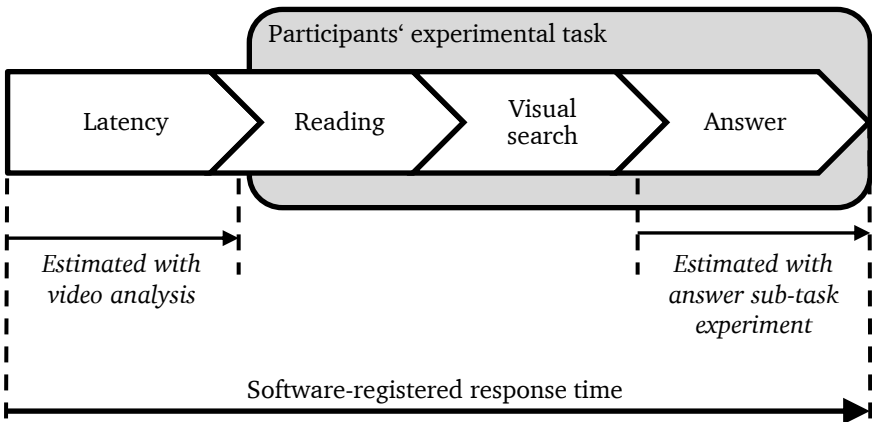


Figure 4.8.: Components of response time. [illustration by author]

By estimating the duration of the latency and the participant's time spent on the sub-task *c*), the reaction time t_R required for sub-tasks *a*) and *b*) was estimated. All 100 software-registered response times per participant were treated as follows: First, mean latencies for the display were subtracted. These were known from video analysis (Table 4.6). Secondly, an estimate for the participants' response time to sub-task *c*) was also subtracted. This second estimate was set to the value of the between-participants mean response time to the answer sub-task experiment of appendix D. Its value varied solely with the participant's given answer cell (see Table D.1a).

All answers having angular error $\Delta\gamma \geq 75^\circ$ or reaction time $t_R \leq 0$ s were considered to be univariate outliers. They were removed, and when necessary, they were conservatively replaced with the participant's average values of the respective dependent variable for that factor level of flight condition.

The assumptions and prerequisites necessary for the following statistical procedures were checked beforehand. Where necessary, corrections were performed. For example, for statistically significant non-sphericity, the Greenhouse-Geisser correction was applied to the relevant degrees of freedom of an analysis of variance. All statistics tests were performed at a significance level $\alpha = .05$, except when stated otherwise.



5 Results on usability and discussion of results

The data collected during the simulator study is analyzed in depth, leading to the results presented in the following chapter. For each of the three hypotheses of section 3.3, inferential statistics are used to analyze the data. Additionally, explorative analyses of the data are performed as well. For easier readability and didactic reasons, the results are discussed directly following each statistical test procedure. At the end of the chapter, the results are discussed with regard to the Global Hypothesis. Also, practical implications for glider flight operations are pointed out.

Throughout the following chapters, it is assumed that readers are familiar with the standard inferential statistics tests of human factors research. Those readers unfamiliar with the subject are pointed to the statistics textbooks of Ross [120], Tabachnick and Fidell [138], Fields [60] and Pallant [114].

Statistical tests with many data points are seldom performed by hand. Most tests in this thesis were performed using computer software, namely IBM SPSS Statistics, Version 21, and the MATLAB statistics toolbox implemented in the releases R2011b and R2014b.

5.1 Visual search error (Hypothesis 1)

The visual search error $\Delta\gamma$ was grouped by unique combinations of different factor levels of flight condition, display format and signal number. These groups exhibited strong non-normality¹ and heteroscedasticity². Applying a square root transformation [113] of the form

$$\Delta\tilde{\gamma} = \sqrt{\frac{\Delta\gamma}{\max \Delta\gamma}} \quad (5.1)$$

improved the distribution characteristics of the transformed visual search error $\Delta\tilde{\gamma}$. In this instance $\max \Delta\gamma = 75^\circ$ was the upper boundary of the visual search error

¹ assessed using 300 Kolmogorov-Smirnov tests for normality. All 300 tests provide statistically significant results at a Holm's step-down adjusted significance level.

² assessed using a Brown-Forsythe test for homogeneity of variances

according to the outlier definition of section 4.6. The transformed variable distributions were still non-normal³ and heteroscedastic,⁴ yet to a much lesser degree than the untransformed distributions. The square root transformation (equation 5.1) increased monotonically in the relevant range of values. Therefore, it retained the relative order between the untransformed and transformed error distributions.

Olson stated that multivariate analysis of variance (MANOVA) procedures are moderately robust to these violations of normality and homoscedasticity [110; 111]. O'Brien and Kaiser even recommended using MANOVA procedures to analyze univariate repeated-measures data violating the assumptions [108]. Following this recommendation, Hypothesis 1 was assessed using a three-way repeated measures MANOVA, including full interaction between all independent variables, with the Pillai-Bartlett trace Λ as the test score of choice [111; 112]. In this case, the transformed visual search error $\Delta\tilde{\gamma}$ was selected as the MANOVA's only dependent variable. The display format used by each participant was the MANOVA's between-subjects independent variable while the flight condition formed a within-subject independent variable. Since measurements were taken 25 times for each unique factor level combination of participant and flight condition, the signal number - ranging from 1 through 25 - was a further within-subject independent variable. The display format's factor levels were the *low-complexity*, *radar-style*, and *perspective* formats. Flight condition factor levels were *straight and level flight*, *horizontal flight with crosswind from left*, *left-hand turning flight*, and *climbing flight*.

5.1.1 Results

The transformed visual search error's magnitude $\Delta\tilde{\gamma}$ was significantly influenced by the display format ($F(2,118) = 55.12, p < .01$). Calculating Bakeman's recommended effect size metric [11] - the generalized η^2 effect size for repeated measures univariate analyses of variance (ANOVAs) (η_G^2) - revealed that this effect was of small strength ($\eta_G^2 = .05$). Similarly, the flight condition had a small and statistically significant effect on $\Delta\tilde{\gamma}$; $\Lambda = 0.73, F(3,116) = 105.54, p < .01, \eta_G^2 = .04$. A statistically significant interacting effect between the display format and the flight state on $\Delta\tilde{\gamma}$ also existed; $\Lambda = 0.30, F(6,234) = 6.85, p < .01$. Its strength, though, was negligible with $\eta_G^2 = .01$. Figure 5.1 illustrates these effects on the untransformed visual search error's magnitude $\Delta\gamma$.

³ Again, all 300 Kolmogorov-Smirnov provide statistically significant results at a Holm's step-down adjusted significance level.

⁴ assessed using multiple Brown-Forsythe tests for homogeneity of variances and follow-up graphical comparisons

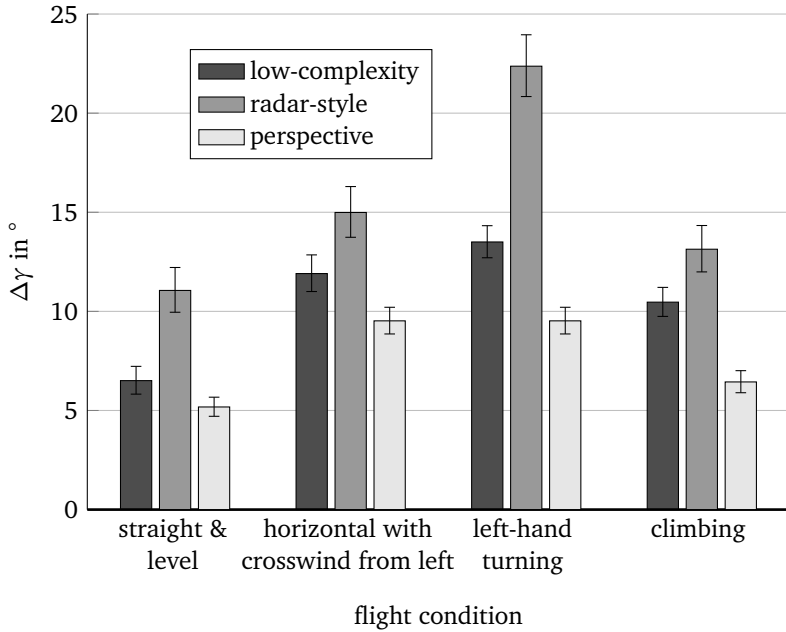


Figure 5.1.: Visual search error magnitude $\Delta\gamma$ (means and 95% confidence interval) grouped by flight condition and display format.

The MANOVA's results revealed differences in the transformed visual search error's magnitude $\Delta\tilde{\gamma}$ between different display format factor levels. Yet, it did not test whether $\Delta\tilde{\gamma}$ was larger for display formats suggesting an ownship-fixed coordinate system (CS) (low-complexity) compared to those suggesting a horizon- and heading-fixed CS (radar-style and perspective). This effect, which was hypothesized in Hypothesis 1, was examined using a pre-planned one-sided first-order contrast [83]. The contrast yielded statistically insignificant results, which was contrary to Hypothesis 1; $F(1,118) = 1.90$, one-tailed $p = .09$. To investigate this unexpected outcome, post-hoc comparisons between $\Delta\tilde{\gamma}$ of the three display format factor levels were performed. Comparisons of the search error for all display format factor levels were statistically significant (see Table 5.1). Their results revealed that the search error of the perspective display format was the lowest, followed by low-complexity format. The radar-style display format had the highest search error. The square-rooted ratio of each comparison's variance compared to the ANOVA's total variance represented the comparison's effect size $r_{\text{effect size}}$ [69].



These comparisons explained between 7% and 39% of the ANOVA’s total variance. This broad range of effect sizes represented small-strength effects⁵ to large-strength effects.⁶

Table 5.1.: Post-hoc analysis: Comparing the difference in transformed visual search error $\Delta\tilde{\gamma}$ between pairs of display format factor levels.

Display format factor levels compared			$F(1,118)$	p	$r_{\text{effect size}}$
Low-complexity	vs.	radar-style	39.48	<.01*	.37
Low-complexity	vs.	perspective	19.08	<.01*	.26
Radar-style	vs.	perspective	110.11	<.01*	.62

* statistically significant at a Holm’s step-down adjusted significance level

Similar post-hoc comparisons were performed for $\Delta\tilde{\gamma}$ at different factor levels of flight condition (Table 5.2). Again, all individual comparisons were statistically significant at an adjusted significance level. However, the effects were all of negligible strength. Nevertheless, the results represented a statistically significant order of increasing visual search error. Visual search error was lowest in the straight and level flight. Next came climbing flight, followed by horizontal flight with crosswind from the left. The largest visual search error was associated with left-hand turning flight. This order was independent of the display format.

5.1.2 Discussion

The results differed from those expected according to Hypothesis 1: The expected interacting effect between the display format and flight condition lacked substantive significance. Nevertheless, an increase in visual search error with increasing deviation from straight and level flight was evident (Table 5.2). This suggests that the relative orientation between the ownship-fixed CS and the horizon- and track-fixed CS influenced the participants’ errors in estimating the traffic’s position. One possible explanation, which will be further explored in chapter 6, is that some participants relied on other CSs than those suggested by the display format when interpreting the display. For these reasons, Hypothesis 1 is only partially rejected.

⁵ in case of the comparison between the low-complexity and perspective display formats

⁶ in case of the comparison between the radar-style and perspective display formats

Table 5.2.: Post-hoc analysis: Comparing the difference in transformed visual search error $\Delta\tilde{\gamma}$ between pairs of flight condition factor levels.

Flight condition factor levels compared					$F(1,118)$	p	$r_{\text{effect size}}$
Straight flight	and level vs.	climbing flight			35.06	$<.01^*$.05
Climbing flight		vs.	horizontal flight with crosswind from left		32.62	$<.01^*$.04
Horizontal flight with crosswind from left		vs.	left-hand turning flight		40.86	$<.01^*$.05

* statistically significant at a Holm's step-down adjusted significance level

Also, the data revealed that the visual search error differed substantively between display formats. As expected, the lowest error was associated with the perspective display format whereas the highest error was unexpectedly made when using the radar-style display. This effect will be analyzed in more detail in the following section 5.2. Nevertheless, a substantive significance of the display format on the effectivity existed. The perspective display format was best suited for guiding the participants' visual attention close to the traffic's actual position in the out-the-window (OTW) view. It was uniformly more robust against misinterpretations than the two commercial off-the-shelf (COTS) display formats studied. The low-complexity display format still provided better effectivity for the task than the radar-style display format. While the radar-style display format is preferred by many highly experienced cross country and contest pilots for its tactical information, it is not well-suited for precisely locating traffic in the OTW view. By more precisely guiding the pilot's gaze towards the direction of expected traffic, the perspective display format is suspected to be superior in supporting the visual perception of traffic. Accordingly, level 1 traffic awareness is expected to increase whenever a perspective display is used by glider pilots.

5.2 Explorative analysis: How different ways of showing distance influence the visual search error

Contrary to Hypothesis 1, the preceding results revealed that the visual search error $\Delta\gamma$ was larger for the radar-style display than for the perspective display. While

comparing the visual representations of both display formats, it became obvious that the visual search error $\Delta\gamma$ was also influenced by the distance to the traffic. To examine whether distance had the same influence on the visual search error $\Delta\gamma$ for both formats, the experiment's data was analyzed exploratively.

A pseudo-linear model depicting the influence of the traffic's slant distance R and the display format's factor level on the transformed visual search error $\Delta\tilde{\gamma}$ was fitted to the data using robust multiple linear regression [104].

$$\Delta\tilde{\gamma} = f_{0,r} + c_p \cdot f_{0,pr} + [f_{1,r} + c_p \cdot f_{1,pr}] \cdot R \quad (5.2)$$

with

$$c_p = \begin{cases} 0 & \text{for radar-style display} \\ 1 & \text{for perspective display} \end{cases} \quad (5.3)$$

In equation 5.2 the parameter $f_{0,r}$ represented the zero-order offset for the radar-style display whereas $f_{0,r} + f_{0,pr}$ was the offset for perspective display. Analogously, the first-order slopes for each display format were described via $f_{1,r}$ and $f_{1,r} + f_{1,pr}$. The dummy variables c_p and $c_p \cdot R$ resulted in heteroscedasticity of residuals⁷ for the model predictors. No multicollinearity existed for the different predictors.⁸ Robust regression methods, relying on bootstrapping, still allowed reliable and unbiased estimators of model parameters to be attained [61, section 8.8].

5.2.1 Results

A multiple linear regression relying on 10^4 bootstrap samples was performed. The forced entry of all four predictor variables created a model with a moderate strength effect size; $\bar{R}^2 = .09$, $F(3,6144) = 203.71$, $p < .01$ with \bar{R}^2 being the ratio of explained variance based on Pearson's adjusted regression coefficient. The estimated parameters of equation 5.2 and their 95% bias-corrected and accelerated confidence intervals are shown in Table 5.3. All parameter estimates were statistically significant at an adjusted significance level. When squaring the respective structure coefficients, it was evident that the modeled parameters contributed between 39% and 67% of the total modeled variance.

Figure 5.2 illustrates how the model fits to the untransformed visual search error distribution $\Delta\gamma$ for both display formats. In the figure, the discrete values of the visual search error, stemming from the experiment's gridded OTW view and touch

⁷ assessed using graphical methods

⁸ assessed using the variance inflation factor

Table 5.3.: Multiple regression results: Parameter estimates with 95% confidence intervals (CIs) and structure coefficients.

Parameter	estimate	95% CI bounds		<i>p</i>	structure coeff.
		lower	upper		
$f_{0,r}$.530	.510	.550	<.01 [*]	
$f_{0,pr}$	-.150	-.172	-.127	<.01 [*]	-.623
$f_{1,r}$	-.051 ¹ /km	-.059 ¹ /km	-.044 ¹ /km	<.01 [*]	-.666
$f_{1,pr}$.015 ¹ /km	.006 ¹ /km	.024 ¹ /km	<.01 [*]	-.818

* statistically significant at a Holm's step-down adjusted significance level

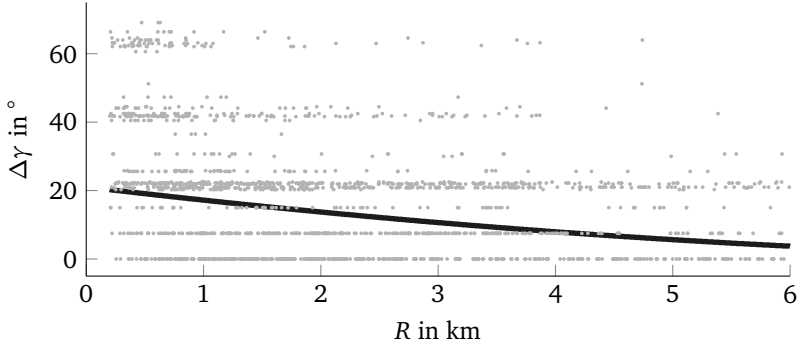
screen monitor response, are evident. Both display formats exhibited decreasing visual search error magnitude with increasing slant distance R . However, the perspective display format consistently provided a lower visual search error than the radar-style display format in the relevant distances between 200 m and 6000 m.

5.2.2 Discussion

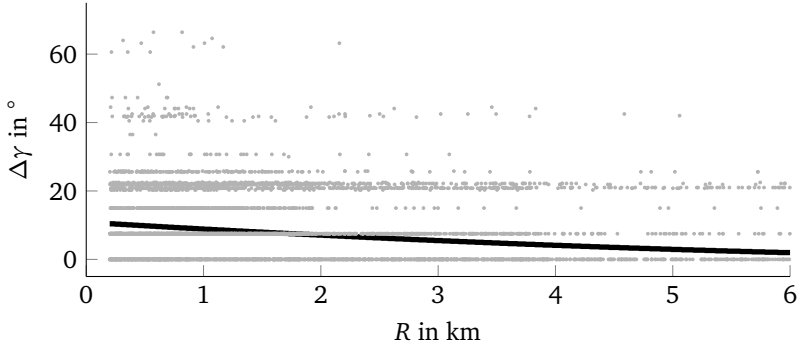
While a correlation between slant distance R and visual search error $\Delta\gamma$ is by no means sufficient to prove causality, it was suspected that the following effects were responsible for the increasing error with decreasing distance. At first, the suspected effect on the radar-style display format will be discussed. The horizontal distance R_H being depicted as a radial coordinate on the radar-style display (compare Figure 3.5). As the traffic's slant distance R decreased its horizontal distance R_H decreased as well.

$$R_H = R \cdot \cos \varepsilon \quad (5.4)$$

This resulted in the traffic symbol moving closer to the ownship symbol at the center of the radar-style display format's two-dimensional polar CS. As it moved closer to the polar CS's origin, judging the traffic symbol's radial coordinate - which represented the traffic's relative bearing ρ - became more difficult. An extreme example of this was when parts of the traffic symbol overlapped the ownship symbol at the center of the polar CS. In this case pilots lacked a reference to judge the traffic's relative bearing ρ . Inversely, it became easier to judge the traffic's relative bearing as the distance to the traffic increased. This led to the visual search error of traffic at the maximum distance $\max R = 6000$ m being an order of magnitude lower than the error associated with traffic being at the minimum distance $\min R = 200$ m.



(a) radar-style display (1670 data points)



(b) perspective display (4478 data points)

Figure 5.2.: Variations of visual search error with distance to traffic: Raw data (dots) and fitted multiple regression model (curved lines).

Despite the fact that the perspective display's visualization differed completely from the radar-style display's, it exhibited a qualitatively similar decrease of the visual search error $\Delta\gamma$ with increasing slant distance R . Due to the differences in visualization, other mechanisms were suspected to be responsible for the correlation of $\Delta\gamma$ and R . As slant distance to the traffic decreased, the size of the traffic symbol and associated text labels grew according to the scaling parameter $k(R)$ of equation 3.2. As it did, the number of pixels used to depict the symbol increased as well. This may have made it more difficult for the pilot to interpret the traffic symbol's coordinates at the center of the glider symbol. This uncertainty may have propagated into uncertain estimates of the traffic's relative bearing ρ and elevation

ε being made by the participant. This, in turn, may have resulted in an increased visual search error.

While the identified multiple regression model suggested the aforementioned mechanisms, it did not prove their existence. Still, the model demonstrated that the perspective display format is superior to the radar-style display format in terms of visual search error over all slant distances studied. The perspective display format was less prone to distance-induced errors in the task of estimating the traffic's direction in the OTW view than the radar-style display format. The fact that the pilot's visual search began at a larger angular distance away from the actual traffic's position is expected to lead to more time being required until traffic is successfully located in the OTW view. In practice, this longer duration is at least partially compensated by increasing traffic conspicuity and traffic visual size as the traffic moves closer [6].

5.3 Reaction time (Hypothesis 2)

Contrary to the distributions of visual search errors in the previous section 5.1, reaction times were normally distributed⁹ and showed only moderate heteroscedasticity.¹⁰ Because ANOVAs are slightly conservative under the given heteroscedasticity, they could still be utilized [158]. Hypothesis 2 was assessed using a univariate three-way repeated measures ANOVA. The ANOVA's independent variables were identical to the MANOVA's of the previous section 5.1. Full interaction between the ANOVA's three independent variables was regarded. Reaction time t_R was selected as the ANOVA's dependent variable.

5.3.1 Results

The flight condition provided a statistically significant influence on reaction time at a negligible strength; $F(3, 354) = 7.37, p < .01, \eta_G^2 < .01$. On the other hand, the display format had no main effect on reaction time; $F(2, 118) = 1.21, p = .30$. However, interaction between flight condition and display format significantly influenced the participants' reaction time; $F(6, 354) = 2.36, p = .03$. This effect was yet again of negligible strength; $\eta_G^2 < .01$. The results are also illustrated in Figure 5.3.

⁹ assessed using 300 Kolmogorov-Smirnov tests for normality. All provide statistically insignificant results at a Holm's step-down adjusted significance level

¹⁰ assessed using a Levene test for homogeneity of variances and graphical follow-up analyses

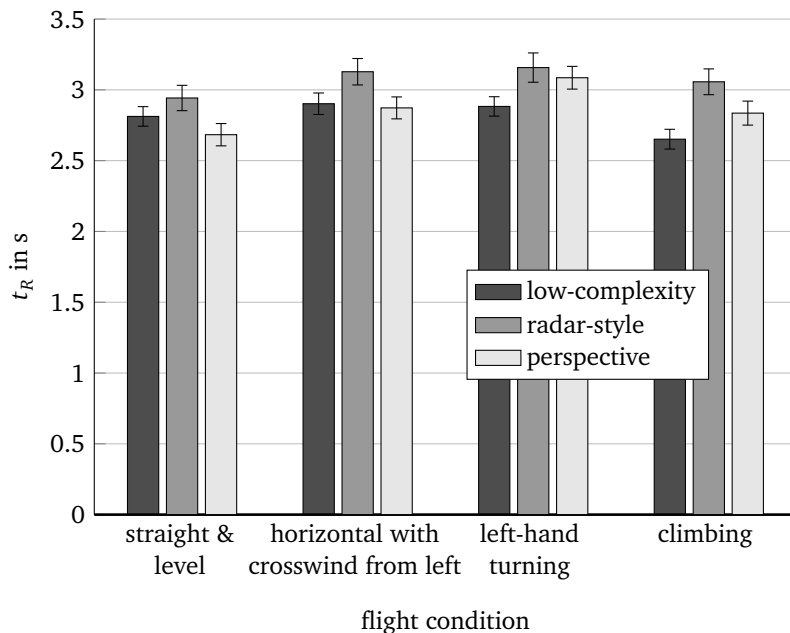


Figure 5.3.: Reaction time t_R (means and 95% confidence interval) grouped by flight condition and display format.

5.3.2 Discussion

Even though statistically significant effects on reaction time existed, those effects were all of negligible strength. Flight condition and display format, as well as their interaction, therefore provided no substantial explanation for the observed variance in reaction time. Thus, Hypothesis 2 was rejected. Mental rotation processes from the suggested CS of a display format to the ownship-fixed CS did not serve to explain the observed differences in reaction time. While pilots likely perform mental rotations to align the display's depictions with the OTW view, these processes are either too swift to make a substantive influence or they don't necessarily take place between the display's suggested and the ownship-fixed CS. The latter option will be explored further in chapter 6.

As already discussed, the temporal cost of using any of the analyzed display formats did not substantively differ during the experiment. Because the temporal cost was quasi constant, effectivity remained as the only influence on the display

formats' efficiency. Thus, the perspective display exhibited the highest efficiency for the task of guiding the pilot's visual attention to traffic in the OTW view. The lowest efficiency was associated with the radar-style display whereas the low-complexity display's efficiency was in-between. Again, the radar-style display exhibited deficits for the task at hand. Yet its strengths for competitive and tactical uses remain.

Several commercially available radar-style displays compensate the reduced efficiency by emulating a low-complexity display when time-critical traffic warnings are issued. In light of the new results, this step is logical. Further, an even higher gain may be expected when future COTS products might revert to showing a perspective display format in time-critical circumstances.

5.4 User satisfaction (Hypothesis 3)

Before testing Hypothesis 3, the distributions of usability and learnability subscale scores of the System Usability Scale (SUS) were examined. Both scores were strongly correlated (Pearson's bivariate correlation coefficient $r = .54$, $p < .01$), which was expected [96]. When grouped by display format, several sub-samples of each score were non-normally distributed.¹¹ Yet for both subscales, homoscedasticity was retained.¹² Hypothesis 3 was evaluated using two separate one-way ANOVAs which are robust against deviations from normality [124]. The significance levels of both ANOVAs were adjusted using Holm's step-down procedure. In both ANOVAs the display format was the independent variable, while one ANOVA's dependent variable was the usability score and the other's dependent variable was the learnability score.

5.4.1 Results

Differences in the SUS subscale scores for both ANOVAs were statistically significant at the adjusted significance level; $F(2, 118) = 3.11$, $p = .05$ for usability and $F(2, 118) = 4.06$, $p = .02$ for learnability. Variations of usability and learnability are illustrated in Figure 5.4. In both cases, the display format had a small effect on the SUS subscale scores ($\eta^2 = .05$ for the usability score and $\eta^2 = .06$ for the learnability score). Planned contrasts between the SUS subscores of the display format

¹¹ assessed using three Kolmogorov-Smirnov tests for normality per subscale score. Their significance levels were adjusted using Holm's step-down procedure. All tests provide statistically significant results, except the test of the learnability subscore's distribution in case of the radar-style display.

¹² assessed using one Brown-Forsythe test for homogeneity of variances per subscale score.

exhibiting high user-centricity in its design process (perspective) and those display formats featuring low user-centricity (low-complexity and radar-style) were performed as well. The perspective display's usability score was significantly higher than that of the remaining displays, with an effect size of moderate strength; $T(118) = -2.25$, one-tailed $p = .01$, $r_{\text{effect size}} = .20$. No such statistically significant difference existed for the learnability score; $T(118) = 0.03$, one-tailed $p = .49$.

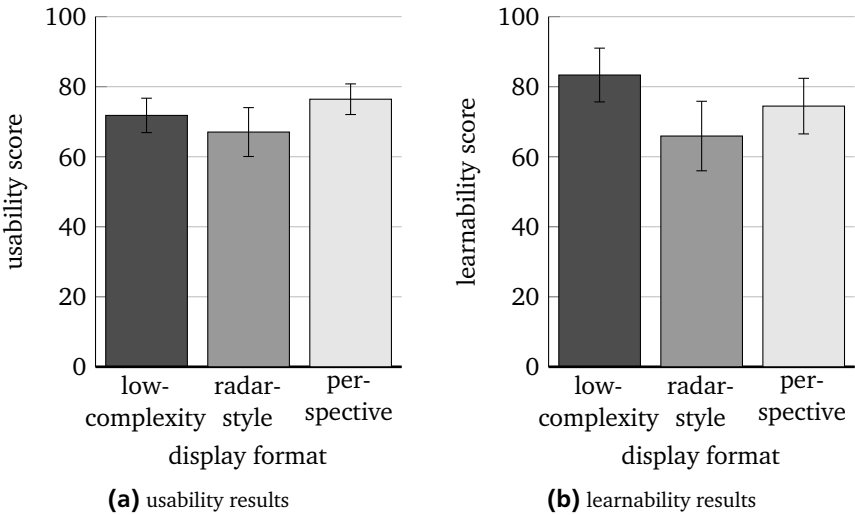


Figure 5.4.: System Usability Scale: Usability and learnability subscale results (means and 95% confidence intervals) grouped by display format.

5.4.2 Discussion

Applying Bangor et al.'s adjective rating scale [12] to the usability score of each display format, participants considered the low-complexity display to be *OK* and bordering *good* for their task. The radar-style display was considered to be merely *OK* while the perspective display was rated to be *good*. Bangor et al. would find the usability subscale score for the low-complexity and perspective displays to be *acceptable* for the given task. The usability score of the radar-style display format would be *marginal high* according to their classification. Nevertheless, participants

felt no major deficits while using any of the analyzed display formats for locating traffic in the OTW view. In case of the usability scores, the statistical results were congruent with those expected according to Hypothesis 3. The data suggested that subjectively perceived usability increased as the degree of user-centricity in the display design process increased.

According to their learnability ratings, participants considered the learnability of the low-complexity and perspective displays to be *good*. On the other hand, they considered learnability of the radar-style display to be *OK*. Bangor et al. would rate these learnability scores as being *marginal high* for the radar-style display and *acceptable* for the remaining formats. However, the statistical results indicated that the subjectively perceived effort required for learning how to use the format did not improve as expected when increasing user-centricity in the design process. The fact that all participants were unfamiliar with the perspective displays format may have contributed to its lower-than-expected learnability ratings. In section 4.2 it was shown that almost all participants were already familiar with the low-complexity display format and some had flight experience with the radar-style display format. Due to its prototypical nature, familiarity with the perspective display format was non-existent. Before deciding whether to retain or reject Hypothesis 3, the role of participants' familiarity with display formats on perceived learnability will be analyzed in more detail.

5.5 Perceived learnability when suppressing familiarity effects

A software design principle states that if a user is more familiar with the design features of a software product, then learnability will be more firmly supported [39, pp. 261–264]. In the study at hand, the participants' familiarity with the different display formats varied notably. Varying familiarity was expected to skew the results of the previous analysis. It was suspected that the participants' learnability ratings of the low-complexity display format were biased to more positive ratings due to their existing familiarity. In contrast, the ratings of the radar-style and perspective formats were suspected to be biased to more negative ratings, caused by a lack of familiarity with the respective display formats.

To suppress this bias effect of familiarity, an additional exploratory analysis was performed. Learnability subscale ratings were reevaluated by purging the ratings of those participants who were already familiar with their display format. As only one participant was unfamiliar with the low-complexity display format, that participant's learnability score was also purged for statistical reasons. Thus, the low-complexity display was excluded from further analysis.

A total of 25 learnability score responses from participants unfamiliar with the radar-style display were retained. Since all participants were unfamiliar with the perspective display all of these responses were also retained. The learnability ratings provided by participants unfamiliar with the radar-style display were normally distributed whereas ratings associated with the perspective display were non-normal.¹³

5.5.1 Results

A non-parametric Mann-Whitney test was performed. The test's independent variable was the display format assigned and its dependent variable was the learnability score provided by each participant. It's results were statistically significant (test score $U = 371.00$, $p = .01$) and can be seen in Figure 5.5. The corresponding rank-biserial correlation coefficient $r = .35$ indicated an effect size of moderate strength.

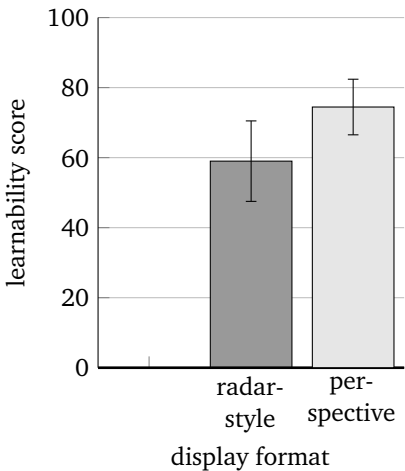


Figure 5.5.: Learnability ratings of participants who are unfamiliar with their display format (means and 95% confidence intervals).

When applying Bangor et al.'s descriptions [12] to the learnability ratings of those users unfamiliar with the radar-style display, learnability of this display

¹³ assessed using two Kolmogorov-Smirnov tests for normality. The significance levels were adjusted using Holm's step-down procedure.

format would be considered to be *marginal low*. The adjective associated with learnability scores remained *good* for the perspective display format.

5.5.2 Discussion

Learning how to use a new display format was perceived differently by participants, depending on the format they faced. Those participants unfamiliar with a given display format perceived the perspective display format as being easier to learn than the radar-style display format. Particularly those participants unfamiliar with the radar-style display format perceived the learning process as being cumbersome when no guidance - such as an expert checkout or a manual - is provided. This difference in perceived learnability effort was attributed to differences in the design features of each display format. These features were a result of the design process. Again, the display format with a higher degree of user-centricity in its design process (perspective) exhibited superior learnability than the display format having experienced a lesser degree of user-centricity in its design (radar-style). Neither the subjective usability results, nor the subjective learnability results, challenged Hypothesis 3. Therefore, the hypothesis was retained.

5.6 Discussion of technology readiness and overall usability

The results of the previous sections clearly showed that the prototype of a perspective display format for a quasi-standard collision alerting system in gliding applications exhibited improved usability in all subjective and objective usability dimensions compared to pre-existing COTS display formats. Since these improvements in usability have been verified in a laboratory environment, the perspective display format's readiness was raised from technology readiness level 3 ("experimental proof of concept" [56, Annex G]) to technology readiness level 4 ("technology validated in lab").

The design process used to create the perspective display prototype distinguished this display format from the two COTS display formats used for comparison. While the COTS low-complexity and radar-style displays were designed using an approach exhibiting a low degree of user-centricity, the perspective display relied on a design process with a high degree of user-centricity. Taking into consideration the perspective display's improved usability traits, it is likely that these traits were a result of the high number of future users involved in the design process. The Global Hypothesis - stating that more user-centric design processes result in optimized usability - was retained.

While the design process of the perspective display format led to a display format which increased pilot performance, the underlying psychological mechanisms were not yet fully understood. Both, the visual search error and reaction time, associated with each display format did not always behave as hypothesized. It is possible that participants not only relied on the display format's suggested CS when interpreting data shown. Instead, their pre-existing knowledge might have resulted in mental models which relied on different CSs. This possibility will be analyzed further in the following chapter 6.

5.7 Criticism of method and potential for future work

Before attempting to link the results and interpretations presented within the preceding sections to practical flight operations, a review of the major weak points of the experiment was appropriate. Both, the experimental method and the prototypical perspective display, had inherent shortcomings.

Since the study was performed as a laboratory experiment, its results are subject to all criticism generic to laboratory experiments. Laboratory experiments only have limited external validity since their environment consists of highly controlled environmental variables whereas the real-world environment does not. The task performed by participants during the experiment, for example, was highly artificial. During actual flight, glider pilots would not only concentrate on interpreting traffic signals, but would need to attend to other simultaneous tasks, such as hand-flying the aircraft, navigating and performing higher-level tactical planning tasks. In real-world flight operations, glider pilots would also receive feedback on whether their search for traffic was successful. This is due to the fact that in real-world flight operations, other traffic is visible OTW. In the simulator, the traffic was not shown in the OTW view in order to prevent participants from adapting their mental models.

Also, the cockpit environment of the experiment was highly unfamiliar to most glider pilots. Multi-seat gliders are most commonly configured with tandem-seat cockpits and conventional instrumentation. The simulator, however, had a side-by-side cockpit and a digital avionics suite installed. This unusual environment may have had influences on the participants' behavior.

The studied flight conditions can also be seen as being artificial. In real-world operations, the pitch, bank and drift angles would vary continuously. In the experiment they were only studied at discrete and fixed values. Only one of these angles deviated from zero during three of the four levels of flight condition. In real-world operations this would be possible only in the calmest of weather conditions.

Perhaps the most artificial aspect in the experiment was the answer task performed by participants. They were asked to estimate the position of traffic indicated on the traffic display via an alphanumeric grid placed on the OTW view. They then provided their estimate on a touch screen monitor. Naturally, no such grid is available in real-world flight operations. Also, the touch screen monitor was only required to record the participants' estimates and would not be available in real-world operations. This method of recording participants' answers was seen as a discretized version of Alexander and Wickens' answer task [2]. The relatively coarse discretization in the experiment may have been responsible for increasing variance in the data observed. Also, it cannot be said with certainty that the answers provided by participants were the actual locations where they suspected traffic to be located.

The experiment still provided insight into the processing of traffic information by glider pilots. Its major advantage, as with any laboratory experiment, was its highly controlled environmental conditions. This increased the experiment's internal validity and allowed for the detection of effects whose effect sizes were previously unknown. It was also the first experiment that systematically analyzed the usability of low-cost collision alerting systems for gliding applications.

Nevertheless, there is room for improvement of the experimental setup. Future work should remove the artificial touch screen response and gridded OTW from the setup. By replacing this with eye-tracking equipment it would be possible to record more naturally the locations where participants would begin their visual search. This would also remove the artificial discretization of the participants' responses. A second area of modification should be the flight conditions. Instead of concentrating on a limited number of pre-selected flight conditions, the pitch, bank and drift angles should be varied continuously. Also, combinations where multiple angles deviate from zero simultaneously should be studied in an attempt to recreate common real-world flight conditions.

Another area of criticism is one of the research objects themselves. Compared to the two COTS traffic display formats, the perspective display format was still in a prototype stage of development. The prototype was optimized for the simulator experiment. It had several design features which might have artificially increased pilot performance. Firstly, only one traffic aircraft could be shown at a time. No methods for preventing clutter or grouping multiple traffic indications were implemented. Additionally, the available physical display area for showing traffic information was larger than the area available on the COTS formats. Whereas the two COTS formats represented traffic in all directions, the perspective display had a limited field of view (FOV). This limited FOV inherently led to a keyhole effect

[159]. Traffic within the perspective display's FOV was well defined. Contrary, traffic outside the FOV could not be shown without some ambiguity.

Future operational versions of a perspective traffic display format for gliding applications would need to overcome the problems of showing traffic outside of the perspective FOV. This may be achieved by integrating the perspective format with an additional traffic representation, such as a radar-style display format. The radar-style display format - contrary to the perspective display format - is also able to show traffic converging from the rear hemisphere. Such multi-display setups, however, may come at the cost of increased mental effort to integrate data shown on both displays. At the same time, the limited FOV of the perspective display already covers most traffic situations. Visual flight rules traffic mainly converges from the frontal hemisphere to the ownship [93]. This majority of traffic, therefore, is within the perspective displays format's FOV. Additionally, aspects such as clutter prevention would need to be addressed.

5.8 Practical implications of results on glider flight operations

Many of the effects identified during the laboratory experiment are expected to carry over into practical glider flight operations. Therefore, guidance for glider pilots using the FLARM low-cost collision alerting system was derived. First and foremost FLARM does not show all relevant traffic. For example, some traffic may not be equipped with a compatible FLARM transceiver or radio signals between FLARM transceivers may be blocked by aircraft structural parts. Thus, glider pilots must still perform their visual scan in order to detect those aircraft not identified by FLARM.

Even if traffic is detected by FLARM, pilots will begin their visual search for the indicated traffic with a random initial offset. These visual search errors increase for all analyzed display formats whenever the ownship's attitude deviates from a straight and level flight condition without crosswind. Whenever the magnitudes of the bank angle $|\Phi|$, the pitch angle $|\Theta|$ or drift angle $|\nu|$ deviates from zero, glider pilots will begin their visual search for traffic in the OTW view farther from the traffic's actual position compared to when they search during wind-free straight and level flight. Such a larger offset in the initial OTW visual search for traffic will lead to a larger volume of airspace which needs to be scanned for traffic. As a result, the time until traffic is visually located is expected to increase. Consequently, glider pilots should plan their tasks in order to allow longer visual searches whenever not flying straight and level.

Particularly prone to a longer visual searches for traffic are flight conditions where large deviations from straight and level flight occur. In gliding such con-

ditions may occur while a) thermaling at high bank angles, b) performing aerobatics¹⁴ with high pitch and bank angles, c) flying in mountain wave conditions with strong winds and corresponding large drift angles, or d) ridge soaring with strong crosswinds and large drift angles in proximity to terrain. All display formats analyzed suffer from reduced pilot performance in these conditions.

¹⁴ Even though forbidden by the FLARM manufacturer [64, p. 12], some glider pilots still use FLARM to help in their visual lookout while performing aerobatics (*name made anonymous*, personal communication, December 2, 2013).



6 Explorative analysis of mental models and their underlying coordinate systems

Results from the effectivity and efficiency analyses showed that an assumption underlying Hypotheses 1 and 2 may have been violated. Apparently, not all participants relied on the coordinate system (CS) suggested by each display format's graphical design features. At least some pilots may have used different CSs while interpreting the traffic information shown to them.

These unexpected results will be exploratively analyzed in this chapter. Suspecting that different participants may have used different personal CSs while interpreting the same display format, several plausible personal CSs are proposed. These CSs are fundamental in describing how each participant may have interpreted the data shown on each display format. Because these CSs describe how participants may have interpreted the data shown, they provide insight into the pilots' mental models (MMs) of the collision alerting system (CAS). Different MMs used by glider pilot participants are identified from the experimental data. Thereafter, the concept of MMs, its influence on pilot performance, as well as its implications for practical glider flight operations, are explored.

6.1 Introduction

While interpreting traffic information shown on the display format, each participant had to define the orientation of a CS in which to interpret the information in. The traffic information then had to be rotated from the participants' ownship-fixed seating position to the CS used for interpretation. Initially, it was assumed that all participants used the CS which was suggested by the design features of the display format they were exposed to. However, the results of the previous chapter suggest that this assumption was violated. Instead, at least some participants may have used different CSs, leading to a personal CS for each participant.

Because different pilots may have different MMs of the CAS, their personal CSs might behave differently to changes in the ownship's attitude. The actual MM used by a participant may have been influenced by the suggested CS of a display format.

Other factors, such as the participant's previous knowledge about the CAS and personal experience with such systems may have influenced their MMs as well.

During the experiment's post-simulator open interview (see section 4.4) many participants were unable to verbalize how they believe that traffic indications on their display format reacted to changes in the ownship's attitude. However, several participants explicitly described that traffic information of the FLARM CAS was always provided in a horizon- and track-fixed CS, due to FLARM's technical limitations. These pilots then mentioned that they consciously neglected to align the traffic information shown on the display with the horizon- and track-fixed CS. They said that aligning the traffic information with an ownship-fixed CS would allow them to begin their visual search for traffic in out-the-window (OTW) more swiftly. According to their explanations, they were aware that by beginning the visual search process in the OTW view more swiftly - through aligning the traffic information with the ownship-fixed CS instead of the heading- and track-fixed CS - a larger visual search error resulted.

Two new hypotheses are derived from the results of chapter 5 and the circumstantial evidence of the open interviews. The first of these hypotheses assumes that the CS suggested by a display format is most important in forming the MMs which glider pilots might develop.

Hypothesis 4: The likelihood with which a MM might be used by a glider pilot when interpreting traffic information depends on the properties of the display format used to present the traffic information. Those MMs that rely on a CS whose behavior is identical to the CS suggested by the display format's design features are the most frequent MMs for glider pilots using that display format.

The second hypothesis postulates that pilot performance varies between different MMs. It is assumed that a trade-off between speed and precision exists.

Hypothesis 5: Different MMs correspond to different pilot performance. If a glider pilot uses a MM which relies on a CS requiring rotation from the pilot's ownship-fixed seating position, that pilot's responses will be more precise, but slower, when compared to pilots using MMs requiring no rotation.

6.2 Method

For further analysis, the MM and associated personal CS of each participant had to be identified. As many participants were unable to verbalize their MMs, they had

to be identified from experimental data available. In the psychological literature, this approach is known as analytical modeling [122, pp. 11–13]. Identifying MMs from the data available was possible because the CSs underlying each MM behaved differently to changes in the ownship's attitude.

6.2.1 An analytical modeling approach to identifying the mental models used by glider pilots when interacting with collision alerting systems

The method for identifying each participant's MM included the following steps:

1. **Propose MMs:** Initially, a finite number of plausible MMs was proposed. Each MM was described by a CS responding differently to variations in the pitch, bank and drift angles.
2. **Predict participants' responses:** The traffic signals shown to each participant were extracted from the data set. Then, each signal was fed into each proposed MM. A prediction was made as to which direction the participant would assume the traffic to be located if she or he were to have exactly this MM.
3. **Assess goodness of fit:** The predicted traffic directions were then compared to the participant's answered directions from the experiment. Differences between predicted and actual answers of a participant were cumulated in a cost function. It was used to compare how well each proposed MM predicted a participant's responses as to where traffic was assumed.
4. **Select best-fitting MM:** The MM whose cost function had the lowest value was the participant's best-fitting MM.

Each of these steps is outlined in more detail below.

6.2.1.1 Proposing plausible mental models

Analytical modeling of MMs relies on plausible MMs to be proposed before the analysis. As mentioned beforehand, each of the proposed MMs was described by a separate personal CS exhibiting unique behavior. This collection of personal CSs was derived using the following train of thought.

Glider pilots are confronted with the concepts of pitch angle, bank angle and drift angle at an early stage in their pre-solo flight training [7, chap. 1; 58, chap. 3].

These concepts should be intuitive to glider pilots and not require further explanation. Coincidentally, the concepts of pitch angle, bank angle and drift angle can be used to describe the relative orientation between FLARM's horizon- and track-fixed CS and the pilot's ownship-fixed seating position (see section A.2.4 for a mathematical description of this coordinate transformation). Each angle describes a rotation that is necessary to align the traffic display's information with FLARM's CS from the pilot's seating position. On a knowledge level, each participant may have known that all of these rotations are required for successful alignment. She or he might also have knowledge deficits, resulting in the participant being unaware of some or all of the required rotations. When experiencing knowledge deficits [133, section 4.2.4] - such as when lacking background information from the FLARM manual - participants may neglect performing some or all of these rotations. In these cases, the personal CS from which the participant may have applied relative bearing and elevation data - as seen on the display - would have been misaligned, and would not have coincided with FLARM's horizon- and track-fixed CS. The disparity between each participant's personal CS and the CS used by FLARM to present traffic information is expected to result in larger visual search errors $\Delta\gamma$.

Mathematically, the transformation from the participant's ownship-fixed seating position was described as a three-step process:

1. Rotation around the \vec{x}_f -axis through the negative bank angle $-\Phi$ if the participant had proper knowledge, or no rotation if a knowledge deficit existed,
2. rotation around the intermediate \vec{y}'''' -axis through the negative pitch angle $-\Theta$ if the participant had proper knowledge, or no rotation if a knowledge deficit existed,
3. rotation around the intermediate \vec{z}'''' -axis through the negative drift angle $-\nu$ if the participant had proper knowledge, or no rotation if a knowledge deficit existed.

This behavior was expressed using identity matrices \mathbf{I}_3 and rotation matrices \mathbf{M} .

$$\begin{aligned} \mathbf{M}_{i,os} = & \left((1 - c_{\nu,i}) \cdot \mathbf{I}_3 + c_{\nu,i} \cdot \mathbf{M}(\vec{z}'''' , -\nu) \right) \cdot \dots \\ & \left((1 - c_{\Theta,i}) \cdot \mathbf{I}_3 + c_{\Theta,i} \cdot \mathbf{M}(\vec{y}'''' , -\Theta) \right) \cdot \dots \\ & \left((1 - c_{\Phi,i}) \cdot \mathbf{I}_3 + c_{\Phi,i} \cdot \mathbf{M}(\vec{x}_{os} , -\Phi) \right) \end{aligned} \quad (6.1)$$

Combining these three steps, the rotation matrix $\mathbf{M}_{i,os}$ described the transformation from the participant's ownship-fixed seating position to the orientation of the CS

of a certain MM i . The variables $c_{\Phi,i}$, $c_{\Theta,i}$ and $c_{\nu,i}$ were binary selector switches to describe whether the respective rotation step was performed ($c = 1$) or ignored ($c = 0$) in the proposed MM. The entries of the rotation matrices \mathbf{M} and identity matrices \mathbf{I}_3 were specified further.

$$\mathbf{M}_{i,os} = \left((1 - c_{\nu,i}) \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + c_{\nu,i} \cdot \begin{bmatrix} \cos \nu & -\sin \nu & 0 \\ \sin \nu & \cos \nu & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \cdot \dots \\ \left((1 - c_{\Theta,i}) \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + c_{\Theta,i} \cdot \begin{bmatrix} \cos \Theta & 0 & \sin \Theta \\ 0 & 1 & 0 \\ -\sin \Theta & 0 & \cos \Theta \end{bmatrix} \right) \cdot \dots \quad (6.2) \\ \left((1 - c_{\Phi,i}) \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + c_{\Phi,i} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \Phi & -\sin \Phi \\ 0 & \sin \Phi & \cos \Phi \end{bmatrix} \right)$$

Since there were three different binary states describing whether the participant was aware of any of the three required rotations, a total of $2^3 = 8$ different mental models were proposed. They are presented in Table 6.1.

Table 6.1.: Characteristics of the proposed mental models (MMs).

MM i	$c_{\Phi,i}$	$c_{\Theta,i}$	$c_{\nu,i}$	comment
1	1	1	1	always horizon- and track-fixed
2	1	1	0	always horizon- and heading-fixed
3	1	0	1	
4	1	0	0	
5	0	1	1	
6	0	1	0	
7	0	0	1	
8	0	0	0	always ownship-fixed

The presented set of plausible MMs was able to reflect the personal CS used by each participant. In practice however, they were not able to show whether the participant used a given MM due to certain knowledge - or lack thereof - or whether the participant consciously elected to use a different MM. Also, it was assumed that each participant's MM remained constant during the course of the experiment. This assumption was based on the fact that the lack of OTW visualization of traffic in the

experiment resulted in a lack of feedback to the participant. This lack of feedback ideally suppressed the learning process during the experiment [133, pp. 71–73], resulting in a constant MM.

6.2.1.2 Predicting participants' responses

When asked where the traffic shown on the CAS display may be located in the OTW view, participants had to base their responses on the relative bearing and elevation stimuli provided on the display. While using the analytical modeling approach to predict each participant's response, the predicted relative bearing ρ^P and the predicted elevation ε^P were mapped relative to the CS of the participant's MM. The traffic location \vec{X}^P predicted by the analytical modeling approach was therefore expressed as

$$\frac{\vec{X}_i^P}{|\vec{X}^P|} = \begin{pmatrix} \cos \varepsilon^P \cdot \cos \rho^P \\ \cos \varepsilon^P \cdot \sin \rho^P \\ -\sin \varepsilon^P \end{pmatrix}_i. \quad (6.3)$$

Transforming the predicted traffic location into ownship-fixed coordinates led to the predicted visual bearing τ^P and predicted visual elevation δ^P .

$$\frac{\vec{X}_{os}^P}{|\vec{X}^P|} = \mathbf{M}_{i,os}^{tr} \cdot \frac{\vec{X}_i^P}{|\vec{X}^P|} = \begin{pmatrix} \cos \delta^P \cdot \cos \tau^P \\ \cos \delta^P \cdot \sin \tau^P \\ -\sin \delta^P \end{pmatrix}_{os} \quad (6.4)$$

The angular difference between the participant's suspected traffic direction \vec{X}^S (as answered during the experiment) and the predicted traffic direction \vec{X}^P of the analytical modeling approach allowed an angular prediction error $\Delta\gamma^P$ to be defined:¹

$$\Delta\gamma^P = \angle(\vec{X}^S, \vec{X}^P) \quad (6.5)$$

How well a given MM predicted the participant's actual responses was achieved using a cost function J . It consisted of the weighted square sum of each participant's angular prediction errors:

$$J = \sum_{\substack{\text{flight} \\ \text{conditions}}} w \sum_{\text{signals}} (\Delta\gamma^P)^2 \quad (6.6)$$

¹ The angular prediction error $\Delta\gamma^P$ was determined analogously to the visual search error $\Delta\gamma$. See equation A.10 in section A.3 for more details.

The cost function included all responses to all signals over all four factor levels of flight condition by a given participant. Large prediction errors incurred a disproportionately higher cost than small prediction errors. Principally, the cost function may be skewed by missing values. A high number of missing values reduces the number of summands, therefore resulting in a lower cost function. Error deflation was compensated by multiplying the square sum of prediction errors for each factor level of flight condition with a pre-specified weight w . If the data contained no missing values for a given factor level of flight condition, then that factor level's weight w was set to $w = 1$. If missing values existed, then the weight w of that factor level was increased proportionately to the number of missing values, thus raising its cost. This way, the cost function did not prefer factor levels of flight condition with high numbers of missing values.

Mapping stimuli to predicted response direction

In practice, the participant had to map the relative bearing and elevation information from the display to the OTW view, using their personal CS. However, the mental map created by this process may have been distorted by numerous psychophysiological effects,² such as the oculogravic illusion [15, p.304]. Also, the map was discretized when only discrete information was shown on the display. Thus, an interrelationship between the stimuli on the display and the predicted relative bearing ρ^p and the predicted elevation ϵ^p of the analytical modeling approach had to be modeled.

This interrelationship varied between the different display formats. In case of the low-complexity display format, participants were exposed only to discrete stimuli, requiring discretized mapping of stimuli to relative bearings and elevations. Contrary, quasi-continuous relative bearing and elevation stimuli were presented on the radar-style and perspective display formats. Thus, continuous mapping functions were selected for these display formats. Simple mathematical functions were chosen to model the mapping process. Following this, the parameters of these functions were identified from the participant's experimental responses.

Mapping of discrete stimuli: In case of the low-complexity display, participants were exposed to a total of five discrete elevation light-emitting diodes (LEDs) and eight discrete bearing LEDs during the course of the experiment. These discrete stimuli are shown in Figure 6.1.

For each factor level combination of participant and proposed MM, the following procedure was performed to map the discrete LEDs to the predicted response directions. Underlying this method was the assumption that illumination of the same elevation LED would result in the same elevation being suspected by the participant.

² E. Groen, personal communication, July 7, 2015

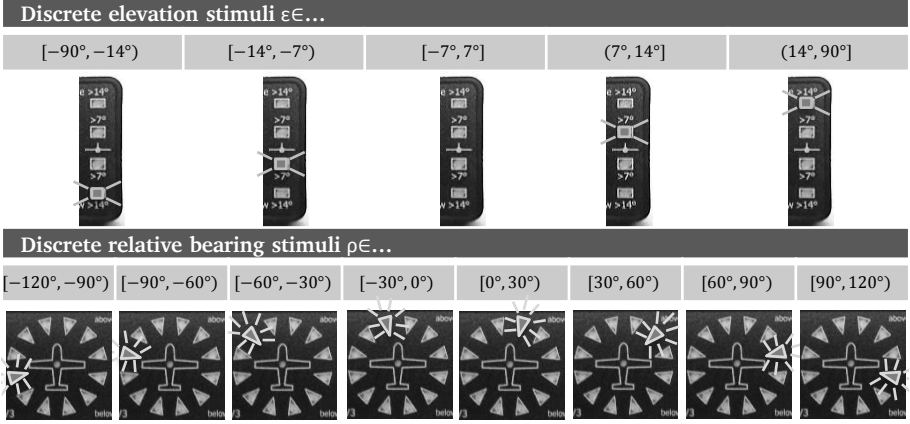


Figure 6.1.: Coding of discrete relative bearing and elevation stimuli of the low-complexity display. [photographs and illustrations by author]

Similarly, illuminating the same relative bearing LED was assumed to result in the same suspected relative bearing. Initially, the participant's responses to the top-most LED stimulus ($\varepsilon \in (14^\circ, 90^\circ]$) were selected. The predicted elevation ε^P for responding to this stimulus was then determined using the Levenberg-Marquardt algorithm for nonlinear optimization [35, section 6.3], so as to minimize the cost function J . This optimization was then repeated for all remaining elevation LEDs and all relative bearing LEDs.

$$J(\text{all discrete stimuli}) = \min J \tag{6.7}$$

This resulted in predicted relative bearings ρ^P and elevations ε^P which minimized the cost function for each individual combination of participant and proposed MM. Essentially, this calibrated the mental map against the participant's experimental data.

Mapping of quasi-continuous stimuli: Contrary to the low-complexity display format, the radar-style and perspective formats showed their relative bearing and elevation information in a quasi-continuous form. Naturally, the way that this information was coded differed between the two display formats (see Table 6.2 for summary). Their discretization effects associated with screen resolution and alphanumeric rounding were small enough to be disregarded.

Table 6.2.: Coding of quasi-continuous relative bearing and elevation stimuli of the radar-style and perspective displays.

Display format	relative bearing ρ	elevation ε
Radar-style	angular coordinate of traffic symbol	implied by horizontal distance R_H (radial coordinate of traffic symbol and alphanumeric value) and relative altitude ΔH (alphanumeric value) ^a
Perspective	horizontal Cartesian coordinate of traffic symbol	vertical Cartesian coordinate of traffic symbol

^a Being provided with values for R_H and ΔH , pilots must mentally estimate the traffic's elevation using the trigonometric function of equation 3.1.

For the two quasi-continuous display formats, the predicted relative bearing and elevation data were mapped proportionately to the indicated relative bearing and elevation.

$$\rho^P = k_\rho \cdot \rho \quad (6.8a)$$

$$\varepsilon^P = k_\varepsilon \cdot \varepsilon \quad (6.8b)$$

The scaling parameters k_ρ and k_ε for mapping the indicated relative bearing ρ and elevation ε from the quasi-continuous display formats to the OTW view were again identified using nonlinear optimization. For each factor level combination of participant and proposed MM, these scaling parameters were selected as to minimize the cost function J .

$$J(k_\rho, k_\varepsilon) = \min J \quad (6.9)$$

As is the case for the mapping of the low-complexity display format, the parameters k_ρ and k_ε calibrated each MM against the participant's experimental data.

6.2.1.3 Assessing the goodness of fit of each mental model

As described in the previous section, the mental maps of all eight proposed MMs were fitted to each participant's experimental responses. This step required determining the cost function J_i for each factor level combination of participant and

MM i . According to equation 6.6, those MMs which predicted the participant's answers well - by having low prediction errors $\Delta\gamma^P$ - also had low cost functions J_i . Contrary, high cost functions corresponded to ill-fitting MMs. Therefore, the cost function J_i quantified how well an MM fit the participant's experimental responses.

6.2.1.4 Selecting the best-fitting mental model for each participant

Now that the goodness of fit was quantified for each combination of proposed MM and participant, the best-fitting MM i with the lowest cost function J_i was selected for each participant. The selected MM i had to satisfy the following equation:

$$J_i = \min_i J_i \quad (6.10)$$

It was assumed that the best-fitting MM of a participant represented the participant's actual MM.

6.2.2 Verification

Before the results from the analytical modeling approach were used for further study, the proposed method had to be verified. The verification procedure is outlined in appendix E. During verification, the models behaved as hypothesized. The analytical modeling approach was therefore considered to be valid.

6.2.3 Treatment of data

Each participant's best-fitting MM was determined using the method above. The reaction time t_R and visual search error $\Delta\gamma$ were taken from the experimental data presented in chapter 4. Also, the same treatment of data as in section 4.6 was applied.

6.3 Results and discussion

Within the following section, hypotheses about the frequency and visual search error magnitudes of different MMs are tested. To simplify reading, each hypothesis is tested and discussed separately. It should be pointed out that, contrary to all other independent variables introduced so far, the best-fitting MM of each participant was not manipulated by the experimenters in the course of the experiment.

Instead, it was merely measured a posteriori. This limits the experiment’s explanatory power to describing statistical correlations between the mental models and the observed dependent variables. Therefore, MMs cannot be identified as being the cause of variations in the data, but only as correlating to these variations.

6.3.1 Frequency of different mental models (Hypothesis 4)

Hypothesis 4 formulated a potential relationship of how often different MMs were expected to occur in glider pilots. It assumed that MMs based on a display format’s suggested CS would be used most frequently by pilots using that display format. In order to test this hypothesis, the frequency of different MMs being identified as best-fitting had to be known. Also, knowledge on how this frequency varied between display formats was required. Absolute frequencies are shown in Table 6.3. Because the number of participants exposed to each display format varied, the absolute frequencies were normalized with the number of participants exposed to the respective display format. The resulting relative frequencies are graphically presented in Figure 6.2.

Table 6.3.: Absolute frequencies of how often different proposed mental models (MMs) were identified as best-fitting.

Display format	proposed MM <i>i</i>								Σ
	1	2	3	4	5	6	7	8	
low-complexity	2	2	2	4	1	2	2	27	42
radar-style	3	4	1	1	2	6	2	14	33
perspective	5	2	0	0	9	17	4	9	46
Σ	10	8	3	5	12	25	8	50	121

In order to test Hypothesis 4, two pre-planned Fisher’s exact tests were performed on different cross tables. The first of these tests examined all data from Table 6.3. According to Hypothesis 4, this test was expected to show that the frequency of MMs varied between different display formats. The second pre-planned test examined the subset of data created by removing the results of the low-complexity display. Because the radar-style and perspective displays both suggested the horizon- and heading-fixed CS, no significant differences in the frequencies of their MMs were expected.

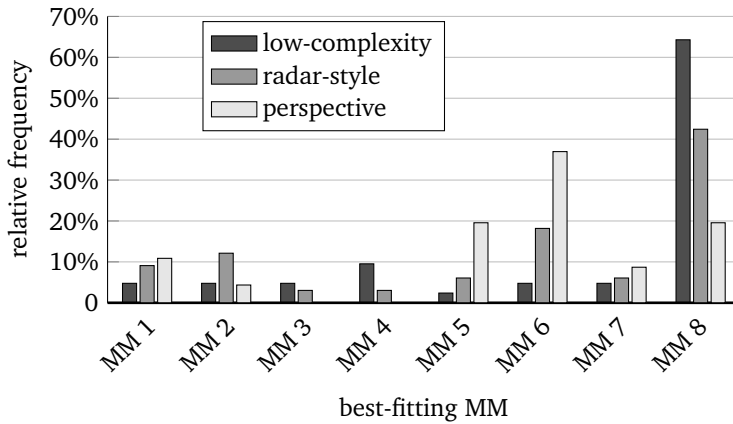


Figure 6.2.: Relative frequencies of how often different proposed mental models (MMs) were identified as best-fitting.

6.3.1.1 Results

The results of the first pre-planned Fisher's exact test - comparing the MM frequencies of all three display formats - showed that there was a statistically significant variation of how often different MMs occurred; $\chi^2(14) = 38.69$, $p < .01$ with a Holm's step-down adjusted significance level. The variation of absolute frequencies between all three display formats was expected according to Hypothesis 4 and it was moderately strong; Cramér's $V = .40$.

Results from the second pre-planned Fisher's exact test performed on the subset of data were not as clear-cut. Because the data from the low-complexity display was removed from the data set, the second pre-planned test only compared the absolute frequencies of the radar-style and perspective displays. The effect bordered statistical significance at an adjusted significance level; $\chi^2(7) = 12.47$, $p = .05$. Yet the effect size corresponded to a moderate-strength effect of varying MMs between the radar-style and perspective display formats; $V = .40$. Upton suggested that these results are substantively significant [149]. Thus, they contradict Hypothesis 4.

The effect that radar-style and perspective displays were associated with different MMs was unexpected. Due to these unexpected results, full pairwise comparison of absolute frequencies between all three display formats was performed post-hoc. This made two additional Fisher's exact tests necessary. The significance values for the post-hoc tests were again adjusted using the Holm's step-down

procedure. The first post-hoc test compared the absolute frequencies of the low-complexity display to those of the radar-style display. It remained statistically insignificant; $\chi^2(7) = 8.55, p = .27$. The low-complexity and perspective display formats were compared in a second post-hoc test, which was statistically significant; $\chi^2(7) = 35.48, p < .01$. The variations of absolute frequencies between those two display formats represented a strong effect; $V = .60$.

Joining the results of the two pre-planned tests and the two post-hoc tests, a clustering of different MMs became evident. The frequencies of best-fitting MMs was similar between the low-complexity and radar-style displays. Both display formats had different distributions compared to perspective display. This contradicted Hypothesis 4. It was assumed that the radar-style and perspective display formats would have similar MMs frequencies, and both would differ from the frequencies associated with the low-complexity display.

6.3.1.2 Discussion

The results of the first pre-planned Fisher's exact test showed that the absolute frequencies of MMs varied between different display formats. The second pre-planned Fisher's exact test compared the absolute frequencies associated with the radar-style and perspective displays. According to Hypothesis 4, it was expected that the second pre-planned test should provide insignificant results, because the radar-style and perspective displays both suggest a horizon- and heading-fixed CS. Instead, it resulted in a p -value barely above the threshold required for statistical significance. At the same time, the associated effect size suggested a moderate-strength effect. Following Upton's guidance for Fisher's exact test, the results were interpreted as being substantively significant [149]. Even though the radar-style and perspective displays suggested the same CS, they did not evoke the same MMs. Hypothesis 4 was rejected due to these results.

Further disproving Hypothesis 4 were the results of the first post-hoc Fisher's exact test. They revealed that the radar-style display had the same frequency distribution of best-fitting MMs as the low-complexity display, although both displays suggested different CSs. MMs relying on the ownship-fixed CS dominated in participants using the low-complexity and radar-style displays. More than half³ of the participants using these display formats relied on this CS by using MM 8.

Revisiting the concept of pictorial realism [119], and how it was applied to the radar-style display format, led to several new findings. While the radar-style display format exhibited design features associated with a horizon- and heading-fixed

³ 41 of 75 participants

CS (MM 2), the display failed to trigger the associated MM in most participants. Possible explanations for this lack of triggering may have been the abstract symbolic and alphanumeric depictions on the display format (compare Figure 3.5). Traffic was depicted using triangles and the ownship was represented by a dot. Geo-fixed features, such as traffic and ownship ground tracks, airspaces, terrain and airfields, were completely missing on the display format. It might be possible that these design features are responsible for not triggering the expected MM in most participants. Instead, most participants used an ownship-fixed CS (MM 8) as their personal CS by default. In case of the low-complexity display format, the static ownship-sketches (compare Figure 3.2) were considered to be responsible for confirming MM 8 and its associated ownship-fixed CS in most participants.

While using the perspective display, participants relied on different MM than the participants working with low-complexity or radar-style displays. This was evident from the results of the second pre-planned and second post-hoc Fisher's exact tests. Only 9 of 46, equaling 19.6%, of participants using the perspective display relied on an ownship-fixed personal CS (MM 8). Conversely, this means that by far most participants performed at least one rotation to align their personal CS. Of the 48 participants using the perspective display format, 33 participants corrected for variations in the pitch angle Θ , 18 corrected for lead angle ν and 7 corrected for bank angle Φ .

The perspective display format's design (compare Figure 3.8) had a well-defined horizon line which remained stationary on the display. At the same time, it lacked an indication that the vertical center line of the display corresponded to the ownship's ground track instead of its heading. According to Hypothesis 4, corrections for the variations in pitch angle Θ and bank angle Φ should have been equally frequent, with only infrequent corrections for the drift angle ν . Why the number of participants correcting for bank angle is much lower than those correcting for pitch angle is not understood. The need for both corrections should have been perceptible according to the concept of pictorial realism [119] from the perspective display format's fixed horizon line. At the same time, the number of participants correcting for the drift angle ν was unexpectedly high, even with a respective clue missing in the display design. Apparently, exposing participants to the novel perspective display format without a ground track clue on the display nevertheless activated passive knowledge in more than one third of the participants. Even without a physical clue, they recognized that a correction for the drift angle was appropriate. Overall, participants using the perspective display made many more of the necessary corrections than participants using the low-complexity radar-style displays.

6.3.2 Pilot performance and different mental models (Hypothesis 5)

Testing Hypothesis 5 was achieved using three pre-planned multivariate analyses of variance (MANOVAs) as well as a series of follow-up contrasts. The initial MANOVAs examined whether participants' performance varied generally between different MMs. Because this was the case, the two performance metrics - visual search error and reaction time - were independently examined using separate MANOVAs. Afterwards, follow-up contrasts were performed on results which were statistically significant according to the MANOVAs.

The analyses had the same independent variables as the analyses of sections 5.1 and 5.3. These were the flight condition and signal number (both within-subject) as well as the display format (between-subjects). In this analysis, the best-fitting MM for each participant was introduced as an additional between-subjects independent variable. Contrary to all other independent variables, the best-fitting MM was a measured factor and was not manipulated during the experiment. Factor levels of all other independent variables were the same as in the preceding sections. The factor levels of the best-fitting MM ranged from MM 1 through MM 8. The MANOVAs regarded full interactions between all independent variables. Dependent variables for the analyses were the transformed visual search error $\Delta\tilde{\gamma}$ and the reaction time t_R .

In previous analyses of similar design (sections 5.1 and 5.3), the prerequisite assumptions of homoscedasticity and normality were non-problematic. Because the same statistical methods as in the preceding sections were used, and these methods have been shown to be adequately robust against violations of assumptions, the assumptions were not examined again.

6.3.2.1 Results

In the initial MANOVA, only the effects associated with the newly introduced independent variable (best-fitting MM) were examined. The two performance metrics - visual search error $\Delta\tilde{\gamma}$ and reaction time t_R - formed the first MANOVA's dependent variables. The best-fitting MM had a statistically significant main effect on the two performance metrics; $\Lambda = 0.31$, $F(14, 198) = 2.65$, $p < .01$. Of the interacting effects studied, only the three-way interaction between the best-fitting MM, the flight condition and the display format exhibited statistical significance; $\Lambda = 0.86$, $F(72, 594) = 1.39$, $p = .02$. Two-way interactions between best-fitting MM and display format - as well as between best-fitting MM and flight condition - did not significantly influence performance ($\Lambda = 0.48$, $F(42, 594) = 1.22$, $p = .16$).

for the 'best-fitting MM * flight condition' interaction; $\Lambda = 0.29$, $F(24, 198) = 1.42$, $p = .10$ for the 'best-fitting MM * display format' interaction).

This allowed for the main effect and three-way interaction on $\Delta\tilde{\gamma}$ and t_R to be studied univariately. Univariate repeated-measures MANOVAs are generally more robust to violations of the normality and homoscedasticity assumptions than traditional repeated-measures analyses of variance [108]. Thus, these univariate analyses were still performed using separate MANOVAs, but with only one dependent variable in each case. The transformed visual search error $\Delta\tilde{\gamma}$ was significantly influenced by the main effect of the best-fitting MM ($F(7,99) = 4.66$, $p < .01$, $\eta_G^2 = .01$) and the three-way 'best-fitting MM * flight condition * display format' interaction ($\Lambda = 0.49$, $F(36,297) = 1.63$, $p = .02$, $\eta_G^2 = .01$). However, both effects were of negligible strength. Figure 6.3 illustrates the untransformed visual search error $\Delta\gamma$ as a function of each participant's best-fitting MM and the flight condition.

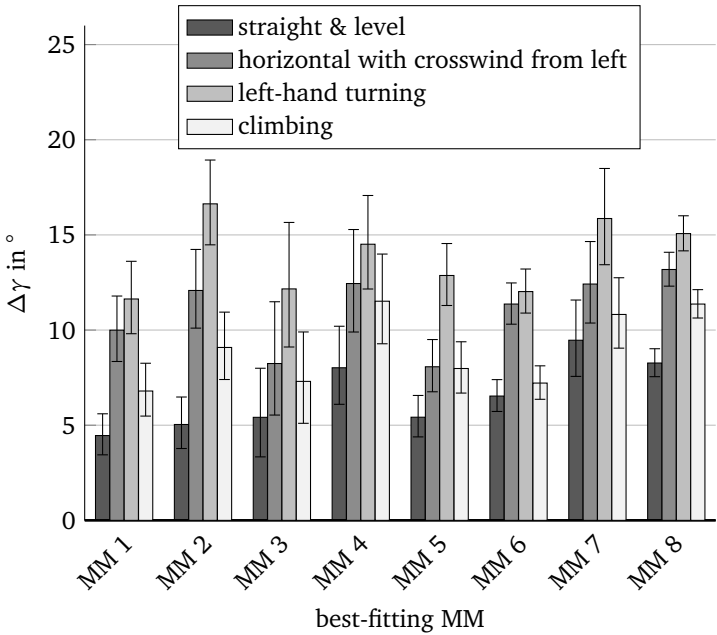


Figure 6.3.: Visual search error magnitude $\Delta\gamma$ (means and 95% confidence intervals) grouped by flight condition and best-fitting mental models (MMs).

Neither the main effect of the best-fitting MM nor the three-way interacting effect between the dependent variables significantly influenced the participants' reaction time; $F(7,99) = 1.91$, $p < .12$ for the main effect; $\Lambda = 0.38$, $F(36,297) = 1.18$, $p = .13$ for the interacting effect.

The effects on the transformed visual search error $\Delta\tilde{\gamma}$ were only statistically significant at minuscule effect sizes. Thus, they lacked substantive significance. Nevertheless, the pre-planned contrasts were performed for academic reasons. For each factor level of flight condition - except straight and level flight - separate contrast analyses were performed. Each contrast analysis compared the transformed visual search error $\Delta\tilde{\gamma}$ of those MMs whose CS coincided with the ownship-fixed CS against those whose CS coincided with the horizon- and track-fixed CS during the selected flight condition factor level. Results from the contrast analyses are given in Table 6.4. Since the CSs of all MMs coincided during straight and level flight, no contrast analysis was performed for this factor level.

Table 6.4.: Pre-planned contrasts: Comparing the difference in transformed visual search error $\Delta\tilde{\gamma}$ between groups of best-fitting mental models (MMs) for different flight condition factor levels.

Flight condition	null hypothesis		$F(1,113)$	one-tailed	$r_{\text{effect size}}$
	$\Delta\tilde{\gamma}(\text{MMs} \dots) \geq \Delta\tilde{\gamma}(\text{MMs} \dots)$			p	
horizontal flight with crosswind from left	1,3,5,7	2,4,6,8	5.77	.01*	.08
left-hand turning flight	1,2,3,4	5,6,7,8	0.15	.35	
climbing flight	1,2,5,6	3,4,7,8	0.80	.02*	.08

* statistically significant at a Holm's step-down adjusted significance level

Two of the three contrasts performed provided statistically significant results at a Holm's step-down adjusted significance level. During horizontal flight with crosswind from the left, the visual search error was significantly smaller for participants which used MMs correcting for the lead angle ν than for participants performing no correction. A similar effect was evident during climbing flight. There, visual search errors for participants using MMs correcting for the pitch angle Θ were significantly smaller than for participants using uncorrected MMs. Both effects, however, were of negligible strength. The third contrast - which concerned left-hand turning flight - was statistically insignificant. During this turning flight, participants using bank-angle-corrected MMs did not show significantly smaller visual search errors than

participants not correcting their personal CS for the change in bank angle. Instead, the corresponding two-tailed p -value ($p=.70$) revealed that the visual search error of the two groups of participants did not differ significantly during left-hand turning flight.

6.3.2.2 Discussion

As mentioned at the beginning of section 6.3, each participant's best-fitting MM was an observed factor. It was not manipulated during the course of the experiment. Therefore, it was only possible to indicate correlations between the best-fitting MM of a participant and her or his performance. It was not possible to derive a causal relationship between pilot performance and the MMs exhibited by glider pilots.

Regarding the results of how pilot performance is influenced by the best-fitting MM, limitations of the MM approach have to be admitted. Seeing that there are no significant effects on reaction time, the concept of MMs does not explain the observed variations in reaction time. These results are similar to those of the usability analysis (section 5.3). In both analyses, it was not possible to detect influences on reaction time. Apparently, the reaction time of participants is influenced by other unknown factors. Because reaction time is a measure of effort, there are no observed differences in effort between glider pilots using different MMs. This fact alone is enough to partially reject Hypothesis 5.

Results of how the visual search error varied between different MMs were almost exclusively conformal to Hypothesis 5. All statistical tests performed on the relationships between a participant's best-fitting MM and her or his exhibited visual search errors were statistically significant and as expected, except for one pre-planned contrast. This confirmed that for most factor levels of flight condition, participants who aligned their personal CS of the orientation FLARM's horizon and track-fixed CS predicted more accurately where traffic was located. The one exception was left-hand turning flight. Here, the results indicated no difference in visual search error for those participants using a bank-angle-corrected MM than for those performing no correction.

Multiple reasons may have led to these indecisive results. The contrast comparing the MMs during left-hand turning flight was a highly unbalanced contrast (see Table 6.4). Only 26 participants relied on MMs correcting for bank angle, while the remaining 95 participants relied on uncorrected MMs. At the same time, Figure 6.3 shows that the confidence intervals associated with left-hand turning flight were generally the largest in the data at hand. Therefore, the mean visual search error of those participants using MMs correcting for variations in bank angle was least precisely known. These facts may have served to mask the hypothesized effect in

the data, making it a statistically insignificant effect instead. Overall, the visual search error data was interpreted as supporting the concept of the MMs. Different glider pilots appear to have different MMs relying on distinct CSs. The orientation of the pilot's personal CS relative to FLARM's horizon- and track-fixed CS influenced the visual search errors of glider pilots. Those pilots aligning their personal CS with FLARM's CS benefited from lower visual search errors.

While the statistical significance of the results served to support the concept of MMs as presented at the beginning of this chapter, the extremely small effect sizes criticized the concept's practicality. Variations in MMs between different participants explain a minuscule part of the variance observed in the overall experiment. Other effects, such as varying the display format or changing the flight condition, were responsible for much larger parts of the observed variance. For all practical purposes in flight operations, the concept of MMs has no substantive significance. While unable to reveal notable variations in pilot performance, it does reveal different mental strategies used by pilots when interpreting the displays. Thus, it is an academic concept which may prove useful during the design and analysis of display formats.

The academic nature of this concept still makes it possible to estimate the effect of the observed MMs on usability. Because it is not known whether certain MMs are causally responsible for a lower visual search error, it is not possible to say whether it is possible to reduce pilots' errors by retraining them on optimized MMs. There is a correlation between a pilot's MM of the CAS and how effectively traffic is located in the OTW view. But for all practical purposes, this correlation is negligible. Following the same train of thought, training pilots to optimize their MMs would not serve to notably alter their reaction time required to interpret traffic information. Efficiency in using a CAS for locating traffic is not influenced by different MMs. If a practical recommendation for glider flight operations can be derived from the observed data, it is thus: Glider pilots should not expend too much conscious effort in attempting to align their personal CS with the one they would suspect to be correct whenever interpreting a traffic display. While flying, glider pilots may notice that they are using an incorrect MM and may consciously attempt to use a correct one instead. Such conscious effort may increase the mental workload without bringing a notable increase in effectivity and efficiency.

In light of the results presented, Hypothesis 5 was partially rejected. While glider pilots do rely on different MMs, and these models are responsible in varying effectivity when using low-cost CASs, these variations are negligible when compared to other effects. Also, there is no effect of these MMs on the time taken to interpret data shown on a traffic display. Therefore, pilot performance does not vary notably

between glider pilots using different MMs. Nevertheless, glider pilots do appear to use different personal CSs. Thus, they have different MMs.

Criticism of the analytical modeling approach and potential for further work

Several shortcomings of the analytical modeling approach may also have been responsible for its lack of operational relevance. For example, it contained the assumption that a participant's MM remained constant during the experiment. Another deficit was that no knowledge on the robustness of the identification method of section 6.2 was available. It still is uncertain whether the method actually detected the MMs used by glider pilots and whether all relevant models were proposed beforehand. Within this thesis only eight different MMs out of a quasi-infinite number of different MMs were proposed. Other MMs might rely on coordinate transformations using other angles besides the pitch, bank and drift angles. At the same time, it is uncertain whether the method robustly identifies the correct MM of a participant, even in light of non-systematic errors being made.

Also, identifying whether the participant performed a rotation around any of the given axes was only achieved through one factor level of flight condition for each of the three axes. It was not evaluated how these discrete combinations of the pitch, bank and drift angles interacted with the robustness of identifying the correct MM of each participant. Future work would need to concentrate on investigating the robustness of the identification method. Since the current research only observed the best-fitting MMs, no causal connection to pilot performance could be deduced from these observations. In order to show such causality, manipulating the MMs of glider pilots would be required. This may be achieved by training them a priori on a certain MM.

6.4 Conclusion

During a comparison study on the usability of different display formats for low-cost CAS (chapters 4 and 5), some unexplained effects on participants' effectivity became apparent. Studying the experimental data post-hoc raised the suspicion that different participants may have interpreted the traffic information on the same display format using different personal CSs. It was also suspected that these personal CSs were fundamental in describing the MMs which participants may have formed while working with one of the displays of the low-cost CAS in the experiment.

From this starting point, a theory of how pilots may use different CSs was developed. Multiple plausible CSs, forming the basis of multiple plausible MMs, were proposed. These MMs were based on potential knowledge deficits which may have

been present in the participants. For each participant, the best-fitting of the pre-proposed CSs - and associated MMs - were identified.

It was then determined how often different MMs occurred. Even though the two display formats exhibited different suggested CSs, the radar-style and low-complexity display formats elicited similar MMs. For both display formats, participants mainly relied on ownship-fixed interpretations of the traffic information. Participants using the novel perspective display relied on different MMs. Even though the perspective display provided a salient horizontal reference which theoretically allowed participants to recognize the need for compensating the pitch and bank angles, most participants only corrected for changes in pitch. Conversely, the perspective display provided in-cockpit clues indicating the need for compensating the drift angle. Nevertheless, some participants were able to activate passive knowledge and correctly compensate for the effects of the drift angle.

The analysis of participant performance during the experiment supported the concept that pilots use different personal CSs as basis of their MMs. Differences in how effectively participants performed their task were explainable using this approach. However, no differences in the temporal cost associated with using different MMs were notable.

The statistical effect of the best-fitting MMs on the participants' task effectivity was of minuscule strength. This showed that the analytical modeling approach of linking MMs to the personal CSs used to interpret traffic information is of little practical importance in flight operations. It is much rather a concept of academic interest relevant during the design and analysis of display formats. Nonetheless, different MMs leave unique traces in participant behavior. Thus, these MMs are identifiable. Even though MMs resulting in different personal CSs had no substantive influence on glider pilot performance when using a CAS, they may have a significant effect in other aviation applications. Thus, future researchers are provided with an additional tool while studying pilot behavior.



7 Summary and conclusions

The dissertation at hand started the first chapter by pointing out the importance of gliding research in an aviation context. Gliding allowed for important technologies to spill over into other areas of aviation in the past. Collision alerting systems (CASs) are one technology which is currently being incubated in the gliding community. The reasons for CASs being developed were explained in chapter 2. There, it was demonstrated that gliders often operate in close proximity to one another. The risk of midair collisions, as being perceived by glider pilots, led to the development of low-cost CASs. They are commercially available and should assist glider pilots in their task of seeing and avoiding other traffic. None of these systems, however, can provide absolute protection against midair collisions. Ambiguous indications on their human-machine interfaces (HMIs) have been cited as possible contributing factors in multiple accidents. Due to this, a need to better understand how glider pilots might interpret or misinterpret traffic information shown on HMIs of low-cost CASs was identified. As this lack of research bordered the field of human factors, several useful constructs of aviation psychology and usability research from literature were introduced.

A market study of existing HMIs for low-cost CASs in gliding was performed in the first half of chapter 3. This market study illustrated how traffic information is currently displayed on commercial off-the-shelf (COTS) products by taxonomizing the HMIs. Taxonomy criteria were the primary sensory channel used to relay information to the pilot, the category of information relayed and the complexity of how this information is presented. Two COTS baseline display formats - having different presentation complexity - were identified for future comparison as well. During the study it was revealed that no COTS product or research application used a perspective presentation of traffic information. According to literature, perspective presentations of traffic information showed potential benefit in military applications. Seeing that the same benefits might also apply to gliding applications, a perspective traffic display was designed using a user-centered design process. This design process was outlined in the second half of chapter 3. A total of 1400 potential users assisted in the process by expressing their preferences in online surveys and participating in the experiments. At the end of that chapter, the characteristics of the newly designed perspective display were compared to those of the two baseline display formats. This led to multiple hypotheses concerning the usability of the three display formats to be compared.

Chapter 4 presented an experimental laboratory setup capable of testing the role played by traffic displays on where in the out-the-window (OTW) view participants suspected traffic to be located. During the experiment, a total of 137 glider pilots were exposed to three different display formats; the low-complexity, radar-style and perspective displays. The experiment was conducted in a “frozen” flight simulator configured in different flight conditions. Namely, these conditions were straight and level flight with and without crosswind, turning flight and climbing flight. Participants were presented with traffic information only on their display, and not in the OTW view. Correspondingly, they provided answers as to where in the OTW view they suspected the traffic to be located.

Results from the preceding laboratory experiment were analyzed using inferential statistics and discussed in chapter 5. Pilot effectivity for the task of estimating traffic position varied between display formats and between flight conditions. The smallest prediction errors were associated with those pilots using the perspective display while the largest errors were made by pilots using the radar-style display. For all three display formats, errors grew as the ownship deviated from straight and level flight. Conversely, reaction time did not vary notably throughout the experiment. Therefore, the highest efficiency was found in pilots using the perspective display format during straight and level flight. Subjective satisfaction was higher for those participants using the perspective display format than those using the two remaining COTS formats. Also, the perspective display’s learnability was perceived to be higher than the radar-style display’s. Overall, the newly designed perspective display performed better in all usability dimensions than the two COTS displays. This proved the perspective display’s potential in gliding applications and allowed its technology readiness level to be raised.

The usability results of chapter 5 suggested that not all participants using one display format may have interpreted the traffic information identically. It was deemed possible that different participants may have used different personal coordinate systems (CSs) when attempting to locate the traffic in the OTW view. This possibility was further explored in chapter 6. Each personal CS was suspected as being the result of each participant’s individual knowledge and to be closely linked to the participant’s mental model (MM) of the CAS. The data gathered in the preceding experiment was analyzed using an analytical modeling approach. Eight MMs and their associated CSs were proposed. The MM best describing each participant’s response behavior during the experiment was identified. The majority of participants using the low-complexity and radar-style displays exhibited a best-fitting MM associated with an ownship-fixed interpretation of the traffic information shown. Thus, they did not correct for any deviations from straight and level flight. The majority of participants using the perspective display format corrected

at least some deviations from straight and level flight by regarding changes in pitch angle, bank angle or drift angle. While the data supported the proposed concept of MMs based on different personal CS, the concept did little to explain differences in performance. For all practical purposes, performance did not vary notably between different participants with different MMs. Though being of little practical significance in actual flight operations, the proposed concept of MMs revealed different strategies used by pilots in interpreting the different display formats. The concept provides a novel tool for analyzing contemporary and novel display designs.

7.1 Recommendations

From the literature review of chapters 2 and 3 and from the experimental results of chapter 5 and section 6.3, multiple recommendations were deduced. All of these recommendations aim at assisting glider pilots in increasing their traffic awareness without experiencing detrimental effects on their remaining aspects of situation awareness or overall performance in flight. Parties who may benefit from these recommendations are glider pilots, aircraft owners and operators, legislative and executive regulatory authorities, glider manufacturers, flight schools and their umbrella organizations, competition rule makers and organizers, CAS designers, as well as HMI designers. Grouped according to the addressed stakeholders, the recommendations are presented below.

Glider pilots

- Glider pilots should familiarize themselves with the operating principles of the CAS installed in the aircraft they fly. They should be able to understand the basic operating principles and limitations of the respective system. Questions worth considering are:
 - How does the CAS work?
 - What information is exchanged between CAS transceivers?
 - How are threat assessments performed?
 - How is traffic of different threat levels indicated?
 - What should be done when receiving a higher-threat traffic alert?
 - What happens if an aircraft close by flies into a 'blind spot' where it is not detected by a CAS?
- Even when having a CAS available, glider pilots should continue to visually scan for traffic. Not all traffic is equipped with a compatible and operational

CAS. There are reports where traffic operated a compatible CAS but effects such as electromagnetic shielding prevented it from being properly received. Therefore, not all traffic will be indicated by the installed CAS.

- Despite traffic being indicated on the HMI of a CAS, it should be identified visually before initiating an avoidance maneuver. Current CASs in gliding neither provide adequate resolution advisories nor adequate information on the traffic's flight state to reliably evade the traffic. At the time, evasion planning can only be reliably performed if the traffic is in sight.
- Whenever attempting to visually identify traffic indicated on the HMI, pilots should take adequate time for this task. They should keep in mind that their initial search will start with an offset from the traffic's actual position. Therefore, it will take time until the traffic is consciously located. The offset and time to find traffic increase as the ownship deviates from straight and level flight. They occur any time that pitch, bank or drift angles are not zero. Finding traffic should take particularly long while a) thermaling, b) performing aerobatics, c) flying in mountain wave conditions with strong winds, or d) ridge soaring with strong crosswinds in proximity to terrain.
- Glider pilots do not need to expend conscious effort in modifying their MM in flight. Thus far, the research has not shown that the initial offset in the visual search for traffic would decrease notably.

Aircraft owners and operators

- Aircraft owners and operators should equip their gliders with a CAS. All commercially available systems are cooperative. Thus, the traffic must be equipped with a compatible transceiver as well. The success of a cooperative CAS requires each aircraft to broadcast its position. Passively receiving this information, without sharing the ownship's position wastes much of the potential benefit of a CAS. Also, aircraft owners and operators should select a CAS which is adequate for the planned flights. FLARM is only one of multiple CASs for general aviation applications. While it is extremely popular in parts of Europe, and forms a quasi-standard for gliding there, this is not the case in other parts of the world. Also, FLARM may not be the system of choice when operating in airspace where high amounts of powered, heavy or fast traffic are expected.
- Aircraft owners and operators should ensure that the CAS is installed properly in their aircraft. Though amateur installations are often permissible, they may result in suboptimal placements of antennas and other problems.

The installation should minimize electromagnetic 'blind spots' - caused by shielding from aircraft structural parts - in which other traffic cannot be reliably received. This may require installations of additional antennas. Operators can confirm the performance of their CAS installation using analysis tools such as FLARM Technology's range analyzer [66].

- Aircraft owners and operators should select an adequate HMI for the CAS installed in their aircraft. The HMI should be adequate for the experience level of the glider pilots flying the aircraft as well as for the flights to be performed. Low-time glider pilots on a local flight in the vicinity of their departure airfield may not have the capacity to interpret the abstract presentations on a radar-style display format and therefore choose to ignore its indications. In contrast, high-time glider pilots on complex cross-country or competition flights may actively seek tactical advantages by using a radar-style display. Of course, the choice of an HMI is influenced by the amount of unused panel space and the availability of glide computers and other avionics, which may serve as non-dedicated displays.

Legislative and executive regulatory authorities

- Legislative and executive regulatory authorities should allow for non-bureaucratic installation and operation of CASs. In Europe, the introduction of low-cost CASs for the gliding community has breached with many conventions of avionics design. Social pressure has led to a swift and widespread proliferation of these systems. Authorities should evaluate whether enhancing existing regulations is the most effective method to achieve increased flight safety. Other mechanisms, such as social mechanisms within the aviation community or financial incentives may allow for a higher acceptance of such measures.
- Legislative and executive regulatory authorities should support the development of an 'integrated' CAS for all airspace users. Current COTS CASs are only isolated solutions for limited groups of operators in the aviation community. FLARM, for example, is primarily used by gliders and light powered aircraft in Europe. It is not used by or able to detect commercial air transports or military aircraft.

Glider manufacturers

- Glider manufacturers should assist aircraft owners and operators in retrofitting existing gliders with CASs. Particularly, recommendations on

optimal antenna placement for each glider type should be made available to aircraft owners and operators.

- In 'clean sheet' glider designs, manufacturers should regard human factors aspects as well as technical aspects of CASs. Traffic indications should be part of the cockpit design philosophy and adequately integrated. If a traffic display is used, manufacturers should make sure that adequate panel space is available in order to prevent cluttering of traffic information.

Flight schools and their umbrella organizations

- Flight schools and their umbrella organizations should integrate popular CASs into their ground school and flight training curricula. Ground school curricula should address the current knowledge deficits in the operating fundamentals of these systems. Flight schools and umbrella organizations should accept the fact that many glider pilots do not actively study the manual of systems installed in their aircraft. Current flight training sessions don't provide glider pilots with guidance on how they should behave in case of traffic alerts. Potential learning objectives to be addressed in these sessions are the following: Glider pilots should become aware that ...
 - low-cost CASs don't provide alerts for all conflicting traffic.
 - low-cost CASs don't provide resolution advisories.
 - visually identifying the traffic is necessary for reliably resolving traffic alerts.
 - even when provided with traffic information or traffic alerts, the visual search for traffic will start with an initial offset. Thus, the search is not an instantaneous process. Instead, it requires time and attention.
 - the initial offset when visually searching for traffic increases whenever deviating from straight and level flight. Thus, the time required for successfully identifying traffic also increases.
 - resolving traffic alerts may draw time and attention away from other high priority tasks certain during phases of flight, such as on final approach.
- Flight schools and their umbrella organizations should not attempt to actively manipulate the MMs of student pilots in practical flying sessions. The work presented does not foresee a notable increase in effectivity when using correct CSs and associated MMs. While not notably increasing effectivity, the pilot's conscious effort may cause workload, negatively impacting pilot efficiency in using the CAS.

Competition rule makers and organizers

- Competition rule makers and organizers should provide competition pilots with incentives to use CAS and similar safety systems. This recommendation was already extended by de Boer [18]. He argued that competition pilots are role models to many other glider pilots. Acceptance of safety systems might rise throughout the community if competition pilots use them. Instead of relying solely on regulation, more subtle social mechanisms - such as incentives by competition rule makers - may also achieve the desired goal of increasing flight safety while at the same time benefiting from increased acceptance within the community.

CAS designers

- CAS designers should assist owners, operators and manufacturers in retrofitting current aircraft with CASs. Together with aircraft manufacturers, they should determine optimal antenna placement for each glider type. Also, they should be able to provide information on practical installations matters, such as what maximum cable length is able to bridge the distance between antenna and transceiver unit without needing signal amplifier.
- CAS designers should make CASs work the way most glider pilots believe them to work. Most glider pilots using COTS display formats, such as the low-complexity and radar-style displays, believe that traffic is indicated in an ownship-fixed polar CS. If traffic information were shown in such an ownship-fixed CS, the CAS would need to be equipped with an attitude and heading reference system (AHRS). A retrofit AHRS solution could receive traffic information from the transceiver using the FLARM data stream [65]. The AHRS would use attitude information about the ownship to transform the traffic information into ownship-fixed coordinates and pass this data on to an HMI. This approach of installing an AHRS “downstream” from the FLARM transceiver would allow AHRSs to be retrofitted into already FLARM-equipped aircraft. If a “clean sheet” approach to a novel CAS is being made, then an AHRS should be integrated at this phase.

HMI designers

- HMI designers should use clearly identifiable and non-abstract symbols to suggest CSs to glider pilots. They should also expect that such symbols may activate passive knowledge in glider pilots about how a system works.
- HMI designers should avoid displaying time-critical traffic alerts on radar-style displays. The visual search error made after retrieving this information

are high and will result in long times until the conflicting traffic is visually identified in an already time-critical situation. Low-complexity and perspective displays result in glider pilots visually identifying traffic more swiftly. HMI designers already switching their displays from a radar-style representation to a low-complexity display in time-critical alerts should continue to do so. Furthermore, they should evaluate whether glider pilots might benefit even more from a perspective presentation of this time-critical information.

- When developing a radar-style display format, HMI designers should provide salient markers of the ownship's position. They should also prevent bearing indications from becoming ambiguous whenever traffic is close by.
- When developing a perspective display format and coding distance information using symbol size, HMI designers should restrict symbols only to reasonable sizes. On the one hand, symbols becoming too large when traffic is close by make it difficult for glider pilots to accurately determine the traffic's position. Small symbols of faraway traffic, on the other hand, make it difficult for glider pilots to detect and discern the symbols on the display.
- For those HMIs which are perceived as being only marginally learnable, HMI designers should assist flight schools and their umbrella organizations in developing training materials for these HMIs. Currently, glider pilots actively seeking additional training on such marginally learnable HMIs are left without formalized guidance.

7.2 Future scientific work

Looking at the research presented, potential for future scientific work was identified in previous sections. Some of this potential is directly related to the limitations of the experimental and analysis methods used.

Potential for improving the experimental setup has been described in section 5.7. The experiment could be optimized by replacing the discrete answer method, realized using the gridded OTW view of the simulator and touch screen monitor, with a continuous answer method. In practice, this could be done by installing multi-camera eye-tracking equipment capable of tracking the participant's eye movement throughout the simulator's $200^{\circ} \times 35^{\circ}$ field of view. Additionally, a future experimental setup might replace the nominally scaled single independent variable of flight condition with three separate ratio scaled independent variables, one each for the pitch angle Θ , bank angle Φ and drift angle ν . By varying multiple angles

non-discretely simultaneously, a much higher amount of practically relevant flight conditions could be analyzed experimentally.

Also, the technology readiness level of the perspective display must be raised further before a commercially viable product can be derived. One topic needing to be addressed is the aspect of depicting traffic outside the perspective display's field of view. Preventing the risk of clutter through symbolic clustering or filtering algorithms also needs to be assessed.

As mentioned while discussing the performance variations of different MMs (section 6.3.2.2), if any statement causally linking different MMs to pilot performance is to be made, a future experimental setup would need to manipulate participants' MMs to demonstrate this causality. Additionally, further knowledge about the procedure used to identify the different MMs is required for reliable statements to be made. Such future work should assess whether all relevant MMs are proposed during analytical modeling. Additional research should demonstrate the robustness of analytical modeling in identifying the correct MM of participants as well. Once these open questions are clarified, the concept of different MMs resulting in different personal CSs which pilots use while interpreting cockpit displays may prove to be a valuable human factors tool.

In a broader sense, the thesis at hand showed that human factors is a discipline which has not yet received major attention in the gliding community. While isolated studies have applied different research methods to gliding use cases, neither the glider cockpit as a workplace, nor the organizational structures found in the gliding community have been analyzed in depth from a human factors perspective. Within this thesis, only the small task of how glider pilots use low-cost CASs to help them locate other traffic has been studied. How this task is achieved by the relatively heterogeneous glider pilot population while performing complex higher-level aviation and sporting tasks for recreational purposes has not yet received attention.

Overall, there is large potential for applying human factors methods to gliding. Due to a lower technical system complexity compared to many other aviation domains, non-aviation researchers will find it easier to apply their expertise to gliding. This new methodical knowledge might then spill over to the remaining aviation domains. Thus, gliding could once again prove itself as "an important incubator" [142] for all of aviation.



References

1. ABOBA Elektronik. (2007). *External display for FLARM - version 3; display V3: Operating manual* (V 5.1d). Walliswil, Switzerland: Author.
2. Alexander, A. L., & Wickens, C. D. (2004). Measuring traffic awareness in an integrated hazard display. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48, 171–175. doi: 10.1177/154193120404800137
3. Amendt, R., Knebel, C., & Wolff, W. (2006). Ophthalmologie [Ophthalmology]. In H. Pongratz (Ed.), *Kompendium der Flugmedizin* [Aeromedical compendium] (revised ed., pp. 119–138). Fürstentfeldbruck, Germany: Surgeon General of the German Air Force, Aeromedical Institute. (Original work published 2004)
4. American Psychological Association. (2007). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author. (Original work published 2001)
5. Andrews, J. W. (1977). *Air-to-air visual acquisition performance with pilot warning instruments (PWI)* (Project Rep. No. ATC-73). Lexington: Massachusetts Institute of Technology, Lincoln Laboratory.
6. Andrews, J. W. (1991). *Unalerted air-to-air visual acquisition* (Project Rep. No. ATC-152). Lexington: Massachusetts Institute of Technology, Lincoln Laboratory.
7. Apel, K.-H. (1996). *Segelflug Praxis: Ausbildung, Weiterbildung, Inübnghaltung* [Gliding practice: Initial training, recurrent training, maintaining currency] (2nd ed.). Heusenstamm, Germany: Deutscher Aero Club Wirtschaftsdienst.
8. Aretz, A. J. (1991). The design of electronic map displays. *Human Factors*, 33, 85–101.
9. Aretz, A. J., & Wickens, C. D. (1992). Mental rotation of map displays. *Human Performance*, 5, 303–328. doi: 10.1207/s15327043hup0504_3
10. Australian Transport Safety Bureau. (2004). *Review of midair collisions involving general aviation aircraft in Australia between 1961 and 2003* (Research Rep. No. B2004/0114). Canberra, ACT, Australia: Author.
11. Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37, 379–384. doi: 10.3758/BF03192707
12. Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24, 574–594. doi: 10.1080/10447310802205776
13. Barraci, N. (2009). *Conflict resolution in autonomous operations area airspace* (Doctoral dissertation). Technische Universität Darmstadt, Darmstadt, Germany.

-
14. Baumgartner, T. I., & Maeder, U. (2009). Trajectory prediction for low-cost collision avoidance systems. *Digital Avionics Systems Conference*, 28, 1.C.5-1–1.C.5-8. doi: 10.1109/DASC.2009.5347567
 15. Benson, A. J. (2006). Spatial orientation in flight. In D. J. Rainford & D. P. Gradwell (Eds.), *Ernsting's aviation medicine* (4th ed., pp. 293–306). London: Hodder Arnold.
 16. Berweger, C. D., & Schuler, L. (2008). Simulationen zur Verbesserung des FLARM-Funkprotokolls [Simulation for improving the FLARM radio protocol]. *segelfliegen*, 6(3), 30–33.
 17. Bevan, N., & MacLeod, M. (1994). Usability measurement in context. *Behaviour & Information Technology*, 13, 132–145. doi: 10.1080/01449299408914592
 18. de Boer, E. (2011). SAFETY PAYS! A proposal to the International Gliding Commission. *Technical Soaring*, 35, 37–47.
 19. Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* [Statistics for the humanities and social sciences] (6th ed.). Heidelberg, Germany: Springer.
 20. Brinkmann, G., & Zacher, H. (1999). *Die Evolution der Segelflugzeuge* [The evolution of sailplanes] (2nd ed.). Bonn, Germany: Bernard & Graefe.
 21. Brüning, G., Hafer, X., & Sachs, G. (with Jurzig, W.). (2006). *Flugleistungen* [Aircraft performance] (3rd ed.). Berlin, Germany: Springer. (Original work published 1993)
 22. Bronstein, I. N., Semendjajew, K. A., Musiol, G., & Mühlig, H. (2008). *Desk-Top Bronstein: Taschenbuch der Mathematik* [DeskTop Bronstein: Handbook of mathematics] (Version 7.0). Haan, Germany: Harri Deutsch. Digitized book on CD-ROM.
 23. Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). London: CRC Press.
 24. Brown, M. B., & Forsythe, A. B. (1974). The ANOVA and multiple comparisons for data with heterogeneous variances. *Biometrics*, 30, 719–724. doi: 10.2307/2529238
 25. Butterfly Avionics. (2013). *Butterfly displays: User- and installation manual* (Document version 1.1). Sandhausen, Germany: Author.
 26. Butterfly Avionics. (2014). *PowerFLARM portable manual* (Version 3.41 (international)). Sandhausen, Germany: Author.
 27. Campbell, R. D., & Bagshaw, M. (2002). *Human performance and limitations in aviation* (3rd ed.). Oxford, United Kingdom: Blackwell Science. doi: 10.1002/9780470774472
 28. CarTFT.com. (n.d.). *CTF840-SH Datenblatt* [CTF840-SH data sheet]. Reutlin-

-
- gen, Germany: Author.
29. Casner, S. M. (2010). General aviation. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (2nd ed., pp. 595–628). Burlington, MA: Elsevier. doi: 10.1016/B978-0-12-374518-7.00019-5
 30. Civil Aviation Safety Team / International Civil Aviation Organization Common Taxonomy Team. (2011). *Aviation occurrence categories: Definitions and usage notes* (Version 4.2).
 31. Cohen, B. H. (2013). *Explaining psychological statistics* (4th ed.). Hoboken, NJ: Wiley.
 32. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
 33. Colvin, K., Dodhia, R., & Dismukes, R. K. (2005). Is pilots' visual scanning adequate to avoid mid-air collisions? *International Symposium on Aviation Psychology*, 13, 104–109.
 34. Curtis, M. T., & Jentsch, F. (2010). Aviation displays. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (2nd ed., pp. 439–478). Burlington, MA: Elsevier. doi: 10.1016/B978-0-12-374518-7.00014-6
 35. Dahmen, W., & Reusken, A. (2008). *Numerik für Ingenieure und Naturwissenschaftler* [Numerics for engineers and natural scientists] (2nd ed.). Berlin, Germany: Springer. doi: 10.1007/978-3-540-76493-9
 36. Dekker, S., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology & Work*, 6, 79–86. doi: 10.1007/s10111-003-0136-9
 37. Dell. (2006). *Dell Precision M90 Mobile Workstation*. Round Rock, TX.
 38. Dilo, J. T. (2013). *Visualization of external air traffic in the Diamond Global Canvas visual system* (Unpublished diploma thesis). Technische Universität Darmstadt, Institute of Flight Systems and Automatic Control, Darmstadt, Germany.
 39. Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2004). *Human-computer interaction* (3rd ed.). Harlow, United Kingdom: Pearson.
 40. Dolba, B., & Vaupel, T. (2009). *German Patent No. DE 20 2009 004 658 U1*. Munich, Germany: German Patent and Trade Mark Office.
 41. Donders, F. C. (1969). On the speed of mental processes (W. G. Koster, Trans.). *Acta Psychologica*, 30, 412–431. (Original work published 1868) doi: 10.1016/0001-6918(69)90065-1
 42. van Doorn, R. R., & de Voogt, A. J. (2007). Glider accidents: An analysis of 143 cases, 2001–2005. *Aviation, Space, and Environmental Medicine*, 78, 26–28.
 43. DSX High Tech. (2007). *DSX systems: The T-Advisor module* (V 2.0). Breganzona, Switzerland: Author.
 44. Durso, F. T., & Alexander, A. L. (2010). Managing workload, performance, and situation awareness in aviation systems. In E. Salas & D. Maurino (Eds.),

-
- Human factors in aviation* (2nd ed., pp. 217–247). Burlington, MA: Elviesier. doi: 10.1016/B978-0-12-374518-7.00008-0
45. EDIATec. (2011). *Betriebshandbuch ECW100 Kollisionswarngerät* [Operating manual ECW100 collision warning device] (V 1.5). Thierachern, Switzerland: Author.
46. Endsley, M. R. (1988). Situation Awareness Global Assessment Technique (SAGAT). *Proceedings of the National Aerospace and Electronics Conference*, 3, 789–795. doi: 10.1109/NAECON.1988.195097
47. Endsley, M. R. (2000). Situation models: An avenue to the modeling of mental models. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44, 61–64. doi: 10.1177/154193120004400117
48. Endsley, M. R., Bolté, B., & Jones, D. G. (2003). *Designing for situation awareness: An approach to user-centered design*. London: Taylor & Francis.
49. Erb, M. (2000). *Die allgemeine Luftfahrt in Deutschland: Eine Analyse im Spannungsfeld von Ökonomie und Ökologie* [General aviation in Germany: An analysis in the area of conflict between economics and ecology] (Doctoral dissertation). Technische Universität Darmstadt, Darmstadt, Germany.
50. European Aviation Safety Agency. (2012a). *AMC and GM to Part 21: Acceptable means of compliance and guidance material for the airworthiness and environmental certification of aircraft and related products, parts and appliances, as well as for the certification of design and production organisations*. Cologne, Germany: Author.
51. European Aviation Safety Agency. (2012b). *EASA MB 04/2012: Agenda item 9c: Roadmap for regulation of GA: Annex 2 - issues to address*. Cologne, Germany: Author.
52. European Aviation Safety Agency. (2014, October). *GA roadmap: towards simpler, lighter, better rules for general aviation*. Cologne, Germany: Author.
53. European Commission. (2011a). Commission Regulation (EU) No 1178/2011 of 3 November 2011 laying down technical requirements and administrative procedures related to civil aviation aircrew pursuant to Regulation (EC) No 216/2008 of the European Parliament and of the Council. *Official Journal of the European Union*(L 311/1).
54. European Commission. (2011b). Commission Regulation (EU) No 1332/2011 of 16 December 2011 laying down common airspace usage requirements and operating procedures for airborne collision avoidance. *Official Journal of the European Union*(L 336/20).
55. European Commission. (2012). Commission Regulation (EU) No 923/2012 of 26 September 2012 laying down the common rules of the air and operational provisions regarding services and procedures in air navigation and amending Im-

-
- plementing Regulation (EU) No 1035/2011 and Regulations (EC) No 1265/2007, (EC) No 1794/2006, (EC) No 730/2006, (EC) No 1033/2006 and (EU) No 255/2010. *Official Journal of the European Union* (L 281/1).
56. European Commission. (2014). *Horizon 2020: Work programme 2014–2015: General annexes* (European Commission Decision C (2014) 4995 of 22 July 2014).
57. European General Aviation Safety Team. (2010). *Collision avoidance: Methods to reduce the risk* (Safety promotion leaflet No. GA 1). Cologne, Germany: European Aviation Safety Agency.
58. Federal Aviation Administration. (2003). *Glider flying handbook* (No. FAA-H-8083-13). Oklahoma City, OK: Author.
59. Federal Aviation Administration. (2014). *Aeronautical information manual: Official guide to basic flight information and ATC procedures* (Change 1). Oklahoma City, OK: Author.
60. Fields, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: SAGE.
61. Fields, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: SAGE.
62. FLARM Experiment Software, version 1.40 [Computer software and manual]. (2014). Darmstadt, Germany: Technische Universität Darmstadt, Institute of Flight Systems and Automatic Control.
63. FLARM Technology. (2008). *FLARM compatibility considerations*. Baar, Switzerland: Author.
64. FLARM Technology. (2011). *Operating manual FLARM collision warning unit: Software version 5.00*. Baar, Switzerland: Author.
65. FLARM Technology. (2012). *Data port specifications: preliminary: Status protocol version 6.00*. Baar, Switzerland: Author.
66. FLARM Technology. (2015). *FLARM range analyzer*. Retrieved September 4, 2015, from <http://flarm.com/support/tools-software/flarm-range-analyzer/>
67. Flühr, H. (2010). *Avionik und Flugsicherungstechnik: Einführung in Kommunikationstechnik, Navigation, Surveillance* [Avionics and air traffic control technology: Introduction to communications technology, navigation and surveillance]. Heidelberg, Germany: Springer. doi: 10.1007/978-3-642-01612-7
68. Fuchs, K.-U., Himmler, K., Lampert, P., & Klaas, U. (2015). *Untersuchungsbericht* [Investigation report] (Final Report No. BFU 3X092-14). Braunschweig, Germany: German Federal Bureau of Aircraft Accident Investigation.
69. Furr, R. M. (2004). Interpreting effect sizes in contrast analysis. *Understanding Statistics*, 3, 1–25. doi: 10.1207/s15328031us0301_1
70. Garrecht Avionik. (2011). *TRX-2000: ADS-B traffic monitor: User manual: Installation manual* (Revision 1.0b). Bingen, Germany: Author.

-
71. German Federal Aviation Office. (2014). *Statistik 2013: Lizenzen Luftfahrt-Bundesamt und Bundesländer* [Statistics 2013: Licenses German Federal Aviation Office and federal states; Data file]. Braunschweig, Germany: Author. Retrieved September 4, 2015, from http://www.lba.de/SharedDocs/Downloads/DE/SBL/SBL3/Statistiken/Statistik_Luftfahrtpersonal_2013.html?nn=700678
 72. German Federal Aviation Office. (2015). *Bestand an Luftfahrzeugen in der Bundesrepublik Deutschland* [Inventory of aircraft in the Federal Republic of Germany]. Braunschweig, Germany: Author. Retrieved September 4, 2015, from http://www.lba.de/DE/Presse_POE/Statistiken/Statistik_Luftfahrzeuge.html?nn=700678
 73. German Federal Bureau of Aircraft Accident Investigation. (n.d.). *ECCAIRS accident database: Release 4*. Retrieved from <http://flugunfall-db.bfu-web.de/e4-web/login.htm>
 74. Gibb, R., Gray, R., & Scharff, L. (2010). *Aviation visual perception: Research, misperception and mishaps*. Farnham, United Kingdom: Ashgate.
 75. Gliding Commission. (2010). *Die Segelflugausbildung: Methodik, Richtlinien und Bestimmungen* [Glider flight training: Methods, guidelines and regulations]. Braunschweig, Germany: German Aeroclub.
 76. Haarhaus, R. (Author). (2014, May 24). Crash am Himmel: Spurensuche nach dem Flugunfall [Crash in the sky: Searching for traces after the aircraft accident] [Television series episode]. In R. Schäfer (Executive Producer), *Exklusiv im Ersten* [Exclusively on the First German Television Channel]. Frankfurt am Main, Germany: Association of Public Broadcasting Corporations in the Federal Republic of Germany.
 77. Haberkorn, T., Koglbauer, I., & Braunstingl, R. (2014). Traffic displays for visual flight indicating track and priority cues. *IEEE Transactions on Human-Machine Systems*, 44, 755–766. doi: 10.1109/THMS.2014.2352496
 78. Haberkorn, T., Koglbauer, I., Braunstingl, R., & Prehofer, B. (2013). Requirements for future collision avoidance systems in visual flight: A human-centered approach. *IEEE Transactions on Human-Machine Systems*, 43, 583–594. doi: 10.1109/THMS.2013.2284784
 79. Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years lateral. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50, 904–908. doi: 10.1177/154193120605000909
 80. Heinbücher, K. (2014). *Konzipierung und Präferenzenerhebung eines egozentrischen Verkehrsdisplays für die allgemeine Luftfahrt* [Conceptual design of an egocentric traffic display for general aviation and determination of preferences by future users] (Unpublished master's thesis). Technische Universität Darmstadt, Institute of Flight Systems and Automatic Control, Darmstadt, Germany.

-
81. Hendriks, F. (1974). Dynamic soaring in shear flow. In *AIAA / MIT / SSA 2nd international symposium on the technology of low speed and motorless flight*. New York: American Institute of Aeronautics and Astronautics. (AIAA Paper No. 74-1003) doi: 10.2514/6.1974-1003
 82. Horvath, H., Appenzeller, K., Büttner, K., Friedemann, J., de Sybourg, O., & Gerber, C. (2008). *Zusammenstoss in der Luft eines Motorflugzeuges Type DV 20 und eines Hubschraubers Type AS 332 am 5. März 2007, um 09:53 Uhr UTC, im Platzbereich des Flugplatzes Zell am See, Salzburg* [Mid air collision of an airplane, type of aircraft DV 20, with a helicopter, type of aircraft AS332, on March 5th 2007 along the traffic pattern of the airport of Zell am See (LOWZ), Austria] (Investigation report No. GZ. BMVIT-85.121/0002-II/BAV/UUB/LF/2008). Vienna, Austria: Austrian Air Accident Investigation Bureau.
 83. Howell, G. T., & Lacroix, G. L. (2012). Decomposing interactions using GLM in combination with the COMPARE, LMATRIX and MMATRIX subcommands in SPSS. *Tutorials in Quantitative Methods for Psychology*, 8, 1–22. Retrieved from <http://www.tqmp.org>
 84. Hupfner, P., & Kuttner, W. (Eds.). (2006). *Witterung und Klima* [Weather and climate] (12th ed.). Wiesbaden, Germany: Teubner.
 85. iDisplay versions 2.4.2 and 3.0.1 [Computer software]. (2013). Stuttgart, Germany: SHAPE.
 86. Institute of Flight Systems and Automatic Control. (2014). *Diamond DA 40-180 simulator*. Retrieved June 15, 2015, from Technische Universität Darmstadt, Institute of Flight Systems and Automatic Control Web site: http://www.fsr.tu-darmstadt.de/forschung_und_dienstleistung/ausstattung/dsimulator.en.jsp
 87. International Civil Aviation Organization. (1989). *Pilot skills to make "look-out" more effective in visual collision avoidance* (Circular No. 213-AN/130). Montreal, Canada: Author.
 88. International Civil Aviation Organization. (2010). *Annex 6 to the convention on International Civil Aviation: Operation of Aircraft: Part I: International commercial air transport – aeroplanes* (9th ed.) (Amendment No. 38). Montreal, Canada: Author.
 89. Janke, J., Lang, M., Feuerle, T., Hankers, R., Ronnenberg, A., Stüben, U., & Wassill, K.-H. (2004). *Abschlußbericht BEKLAS: Erkennbarkeit von Segelflugzeugen und kleinen motorisierten Luftfahrzeugen* [BEKLAS final report: Conspicuity of gliders and light powered aircraft] (Research Rep. No. L-6/2002-50.0300/2002). Bonn, Germany: German Federal Ministry of Transportation, Building and Housing.
 90. Jarvis, S. J. (2009). *A misjudged approach to a high accident rate: Exploration of accident causes and instructor decisions relating to inexperienced glider pilots*

-
- (Doctoral dissertation). Cranfield University, Cranfield, United Kingdom.
91. König, C. (2012). *Analyse und Anwendung eines menschenzentrierten Gestaltungsprozesses zur Entwicklung von Human-Machine-Interfaces im Arbeitskontext am Beispiel Flugsicherung* [Analysis and application of a human-centered design process for developing human-machine-interfaces in a working context with exemplary application to air traffic control] (Doctoral dissertation). Technische Universität Darmstadt, Darmstadt, Germany.
92. Knoll, R., Koss, R., & Brandes, T. (2011). *Untersuchungsbericht* [Investigation report] (Final Report No. BFU 3X035-10). Braunschweig, Germany: German Federal Bureau of Aircraft Accident Investigation.
93. Kochenderfer, M. J., Espindle, L. P., Edwards, M. W. M., Kuchar, J. K., & Griffith, J. D. (2009). Airspace encounter models for conventional and unconventional aircraft. *USA/Europe Air Traffic Management Research and Development Seminar*(8).
94. Konrad, G., Colquhoun, W. M., Prô, B., Engels, J., & Feyereisen, T. (2012). Development, flight test and certification of a track-up synthetic vision system with HUD-like symbology for the Pilatus PC-12 NG. *Avionics Europe Conference & Exhibition*.
95. Laine, J., Aaltonen, I., & Halonen, H. (2012). *Collision between two sailplanes in Hattula on 12 June 2011* (Investigation Report No. B1/2011L). Helsinki, Finland: Safety Investigation Authority of Finland.
96. Lewis, J. R., & Sauro, J. (2009). The factor structure of the System Usability Scale. In M. Kurosu (Ed.), *Human centered design* (pp. 94–103). Berlin, Germany: Springer. doi: 10.1007/978-3-642-02806-9_12
97. LG Electronics Deutschland. (2013). *LG D605 Optimus L9II Datenblatt* [LG D605 Optimus L9II data sheet]. Ratingen, Germany.
98. LXNAV. (2014). *LX90xx: GPS navigation system with variometer (including LX9000 and LX9070)* (Version 4.0). Celje, Slovenia: Author.
99. LX navigation. (2014). *LX Flarm color display II: User manual*. Celje, Slovenia: Author.
100. Mehringskötter, S. (2013). *Mentale Modelle beim Umgang mit neuartigen Kollisionswarnsystemen im Luftsport* [Mental models while operating low-cost collision alerting systems in sport aviation] (Unpublished bachelor's thesis). Technische Universität Darmstadt, Institute of Flight Systems and Automatic Control, Darmstadt, Germany.
101. Möller, C., Kostka, A., Neujahr, H., & Klingauf, U. (2012). Enhancing situational awareness by perspective panoramic information presentation. *Proceedings of the European Association for Aviation Psychology Conference*, 30, 232–239.
102. Morgenstern, C. (2012). *Sicherheitskultur und menschliche Fehler im Segelflug*

-
- [Safety culture and human errors in gliding] (Unpublished diploma thesis). Universität Regensburg, Regensburg, Germany.
103. Morris, C. C. (2005). Midair collisions: Limitations of the see-and-avoid concept in civil aviation. *Aviation, Space, and Environmental Medicine*, 76, 357–365.
 104. Nathans, L. L., Oswald, F. L., & Nimon, K. (2012). Interpreting multiple linear regression: A guidebook of variable importance. *Practical Assessment, Research & Evaluation*, 17(9).
 105. National Transportation Safety Board. (n.d.). *NTSB docket management system*. Washington, DC: Author. Retrieved from <http://www.nts.gov/investigations/dms.html>
 106. Naviter. (2013). *SeeYou Mobile help for version 4.50*. Kranj, Slovenia: Author.
 107. Neis, S. (2013). "Midair Collision" trotz Kollisionswarnsystem: Was läuft schief? [Midair collision even with a collision alerting system: What is going wrong?] (Unpublished research seminar report). Technische Universität Darmstadt, Institute of Flight Systems and Automatic Control, Darnstadt, Germany.
 108. O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin*, 97, 316–333. doi: 10.1037/0033-2909.97.2.316
 109. O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance. Volume II, cognitive processes and performance* (pp. 42-1–42-49). New York: Wiley.
 110. Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association*, 69, 894–908. doi: 10.2307/2286159
 111. Olson, C. L. (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, 83, 579–586. doi: 10.1037/0033-2909.83.4.579
 112. Olson, C. L. (1979). Practical considerations in choosing a MANOVA test statistic: A rejoinder to Stevens. *Psychological Bulletin*, 86, 1350–1352. doi: 10.1037/0033-2909.86.6.1350
 113. Osborne, J. W. (2002). Notes on the use of data transformations. *Practical Assessment, Research & Evaluation*, 8(6).
 114. Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS* (4th ed.). New York: McGraw-Hill.
 115. Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision*

-
- Making*, 2, 140–160. doi: 10.1518/155534308X284417
116. Porod, K. (2002). *German Patent No. DE 201 17 525 U1*. Munich, Germany: German Patent and Trade Mark Office.
117. Prinzel, L. J., III, Comstock, J. R., Jr., Etherington, T., French, G. A., Snow, M. P., Endsley, M. R., . . . Corker, K. M. (2004). Human factors issues in synthetic vision displays: Government, academic, military, and industry perspectives. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48, 75–79. doi: 10.1177/154193120404800117
118. Reichmann, H. (2005). *Streckensegelflug* [Cross-country soaring] (10th ed.). Stuttgart, Germany: Motorbuch.
119. Roscoe, S. N. (1968). Airborne displays for flight and navigation. *Human Factors*, 10, 321–332.
120. Ross, S. M. (2006). *Statistik für Ingenieure und Naturwissenschaftler* [Introduction to probability and statistics for engineers and scientists] (3rd ed.; C. Heinisch, Trans.). Munich, Germany: Elsevier. (Original work published 2004)
121. Rouaux, J.-E. (2012). *CIR12_295*. Paris, France: French Gliding Federation.
122. Rouse, W. B., & Morris, N. M. (1985). *On looking into the black box: Prospects and limits in the search for mental models* (Rep. No. 85-2). Atlanta: Georgia Institute of Technology, Center for Human-Machine Systems Research.
123. Scherer, K. (2006, May). Eine Erfolgsgeschichte voller Idealismus: Die FLARM-Biographie [A success story full of idealism: The FLARM biography]. *segelfliegen*, 4, 25–27.
124. Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Buehner, M. (2010). Is it really robust? reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology*, 6, 147–151. doi: 10.1027/1614-2241/a000016
125. Schochlow, V. (2012). *Spatial processing of traffic information in sport aviation - a randomized controlled study of mental rotation processes in non-pilot participants* (Unpublished bachelor's thesis). Technische Universität Darmstadt, Institute of Psychology, Darmstadt, Germany.
126. Schochlow, V., & Gerber, P. (2012). [Literature research on projection formats and display design parameters]. Unpublished raw data. Darmstadt, Germany: Technische Universität Darmstadt, Institute of Psychology, Work and Engineering Psychology Research Group.
127. Schuler, L. (2007). Die Wirksamkeit von Warnsystemen durch Simulation erhöhen [Increasing the effectivity of warning systems through simulation]. *Forum für Wissen: Warnungen bei aussergewöhnlichen Naturereignissen* [Knowledge forum: Warnings of exceptional natural phenomena], 93–96.

-
128. Seth, A., Carlson, K. D., Hatfield, D. E., & Lan, H.-W. (2009). So what? Beyond statistical significance to substantive significance in strategy research. In D. D. Bergh & D. J. Ketchen Jr. (Eds.), *Research methodology in strategy and management* (Vol. 5, pp. 3–27). Bingley, United Kingdom: Emerald.
 129. Shepard, R. N., & Metzler, J. (1971, February). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–703. doi: 10.1126/science.171.3972.701
 130. Shuch, H. (1992). The influence of flight experience on midair collision risk perception. *Accident Analysis & Prevention*, 24, 655–660. doi: 10.1016/0001-4575(92)90017-D
 131. Standards Committee Ergonomics and Standards Committee Information Technology and selected IT Applications. (1999). *DIN EN ISO 9241-11: Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten: Teil 11: Anforderungen an die Gebrauchstauglichkeit - Leitsätze* [Ergonomics requirements for office work with visual display terminals (VDTs): Part 11: Guidance on usability]. Berlin, Germany: Beuth.
 132. Standards Committee Ergonomics and Standards Committee Information Technology and selected IT Applications. (2011). *DIN EN ISO 9241-210: Ergonomie der Mensch-System-Interaktion: Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme* [Ergonomics of human-system interaction: Part 210: Human-centered design for interactive systems]. Berlin, Germany: Beuth.
 133. Steininger, K. (2000). *Handbuch der Flugpädagogik: Lerne Fliegen Lehren* [Handbook of flight educational science: Learn to teach flying] (2nd ed.). Hamburg, Germany: Author.
 134. Steinmetz, L., & Gerber, P. (2013). [Survey on proliferation of low-cost collision alerting systems]. Unpublished raw data. Darmstadt, Germany: Technische Universität Darmstadt, Institute of Psychology, Work and Engineering Psychology Research Group.
 135. Sternberg, S. (1969a). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276–315. doi: 10.1016/0001-6918(69)90055-9
 136. Sternberg, S. (1969b). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 57, 421–457.
 137. Swiss Civil Aviation Safety Officer. (2010). Stellungnahme CASO (letzte Aufdatierung vom 23.03.2010) [Statement of CASO (revised last on May 21, 2010)]. In *Sicherheitsempfehlungen des Büros für Flugunfalluntersuchungen und Stellungnahmen zu deren Umsetzung* [Safety recommendations by the Swiss Aircraft Accident Investigation Bureau and statements about their implementation]

-
- (Rep. No. 2086) (pp. 102–103). Payerne, Switzerland: Swiss Aircraft Accident Investigation Bureau.
138. Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.
139. Le Tallec, C., & Gravier, B. (2002). *European Patent No. EP 1095362 (B1)*. Munich, Germany: European Patent Office.
140. Taneja, N., & Wiegmann, D. A. (2001). Analysis of mid-air collisions in civil aviation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 45, 153–156. doi: 10.1177/154193120104500233
141. Thomas, J. R., Nelson, J. K., & Thomas, K. T. (1999). A generalized rank-order method for nonparametric analysis of data from exercise science: A tutorial. *Research Quarterly for Exercise and Sport*, 70, 11–23. doi: 10.1080/02701367.1999.10607726
142. Thorbeck, J. (2009). Gliding research – an important incubator for the entire aviation. *Technical Soaring*, 33, 7–16.
143. triadis engineering. (2006a). *Altair: Glide computer and navigation system: User manual* (Revision 1.01 EN). Messen, Switzerland: Author.
144. triadis engineering. (2006b). *FLOICE: Kollisionswarnsystem mit akustischer Sprachausgabe: Pilotenandbuch: FLOICE 256: Model A, B und C* [FLOICE: Collision warning system with acoustic voice output: Pilot's manual: FLOICE 256: Model A, B and C] (Version 0.1). Messen, Switzerland: Author.
145. triadis engineering. (2006c). *Vega: Digital voice variometer system: Pilot's manual* (Revision 1.01 EN). Messen, Switzerland: Author.
146. triadis engineering. (2008). *TR-DVS: Pilot's manual* (Document Version 1.1). Messen, Switzerland: Author.
147. Tullis, T. S. (1981). An evaluation of alphanumeric, graphic, and color information displays. *Human Factors*, 23, 541–550.
148. Uhlarik, J., & Comerford, D. A. (2002). *A review of situation awareness literature relevant to pilot surveillance functions* (Final Rep. No. DOT/FAA/AM-02/3). Washington, DC: Federal Aviation Administration, Office of Aerospace Medicine.
149. Upton, G. J. G. (1992). Fisher's exact test. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 155, 395–402. doi: 10.2307/2982890
150. von Viebahn, H. (1997). *Konzeption und Untersuchung eines bordautonomen Systems zur Vermeidung von Kollisionen im Luftverkehr* [Conceptual design and analysis of an autonomous onboard system for collision avoidance in aviation] (Doctoral dissertation). Technische Hochschule Darmstadt, Darmstadt, Germany.
151. de Voogt, A. J., & van Doorn, R. R. (2006). Midair collisions in U.S. civil aviation 2000-2004: The roles of radio communications and altitude. *Aviation*,

-
- Space, and Environmental Medicine*, 77, 1252–1255.
152. Welch, J. F., Bjork, L., & Bjork, L. (Eds.). (1999). *Van Sickles's modern airman-ship* (8th ed.). New York: McGraw-Hill. doi: 10.1036/0070696330
 153. Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current Directions in Psychological Science*, 11, 128–133. doi: 10.1111/1467-8721.00184
 154. Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, 50, 449–455. doi: 10.1518/001872008X288394
 155. Wickens, C. D., Lee, J. D., Liu, Y., & Gordon Becker, S. E. (2004). *An introduction to human factors engineering* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
 156. Wickens, C. D., & Prevett, T. T. (1995). Exploring the dimensions of egocentricity in aircraft navigation displays. *Journal of Experimental Psychology: Applied*, 1, 110–135. doi: 10.1037/1076-898X.1.2.110
 157. Wickens, C. D., Xu, X., Helleberg, J. R., Carbonari, R., & Marsh, R. (2000). *The allocation of visual attention for aircraft traffic monitoring and avoidance: Baseline measures and implications for freeflight* (Tech. Rep. No. ARL-00-2/FAA-00-2). Savoy: University of Illinois at Urbana-Champaign, Institute of Aviation, Aviation Research Lab.
 158. Wilcox, R. R., Charlin, V. L., & Thompson, K. L. (1986). New Monte Carlo results on the robustness of the ANOVA F, W and F* statistics. *Communications in Statistics - Simulation and Computation*, 15(4), 933–943. doi: 10.1080/03610918608812553
 159. Woods, D. D. (1984). Visual momentum: a concept to improve the cognitive coupling of person and computer. *International Journal of Man-Machine Studies*, 21, 229–244. doi: 10.1016/S0020-7373(84)80043-7
 160. XCSOar project. (2015). *XCSOar 6.7: The open-source glide computer: User manual* (For XCSOar version 6.7.7).
 161. Zoon Flight Systems. (2008, September). *Owner's manual: PCAS MRX*. Addison, TX: Author.
 162. Zoon Flight Systems. (2011). *Owner's manual: PCAS XRX* (Manual Version 3.0). Addison, TX: Author.





List of figures

- 1.1 Structure of main part of this thesis. 3

- 2.1 High local traffic density underneath a cumulus cloud during the
32nd FAI World Gliding Championships of 2012 in Uvalde, TX. 7
- 2.2 FLARM transceiver box of hardware version 3. 11
- 2.3 Conceptual sketch of the FLARM system. 12
- 2.4 Midair collisions involving gliders and motorgliders. 15
- 2.5 Levels of traffic awareness and their interaction with a mental model. 19
- 2.6 Dimensions of usability. 20
- 2.7 Transfer of research methods in the thesis at hand. 22

- 3.1 Taxonomy for categorizing human-machine interfaces of low-cost
collision alerting systems. 24
- 3.2 Exemplary low-complexity display: The external FLARM display V3. 28
- 3.3 Visualization of the navigation analogy. 29
- 3.4 Orientation of Cartesian and polar coordinate systems while maneu-
vering. 30
- 3.5 Exemplary radar-style display: The external Butterfly display. 33
- 3.6 Schematic depiction of an alphanumeric display. 34
- 3.7 Structure of the perspective display format’s design process. 39
- 3.8 Perspective display format. 42
- 3.9 Hypotheses and corresponding usability dimensions. 45

- 4.1 Flight simulator cockpit mock-up with gridded projection screen. . . 54
- 4.2 Touch screen monitor with answer grid. 55
- 4.3 Location of the installed displays in the flight simulator cockpit. . . . 55
- 4.4 Computer hardware and interface setup. 56
- 4.5 Flow diagram of the experiment’s sessions. 58
- 4.6 Flow diagram of the experimental procedure. 59
- 4.7 Flow diagram of the participants’ task. 61
- 4.8 Components of response time. 62

5.1	Visual search error magnitude $\Delta\gamma$ grouped by flight condition and display format.	67
5.2	Variations of visual search error with distance to traffic: Raw data and fitted multiple regression model.	72
5.3	Reaction time t_R grouped by flight condition and display format. . . .	74
5.4	System Usability Scale: Usability and learnability subscale results grouped by display format.	76
5.5	Learnability ratings of participants who are unfamiliar with their display format.	78
6.1	Coding of discrete relative bearing and elevation stimuli of the low-complexity display.	92
6.2	Relative frequencies of how often different mental models were identified as best-fitting.	96
6.3	Visual search error magnitude $\Delta\gamma$ grouped by flight condition and best-fitting mental models.	100
A.1	Sequence of transformation steps between coordinate systems. . . .	138
D.1	Hardware setup of the answer sub-task experiment.	151
D.2	Experimental distributions of actual and limited error success rates during answer sub-task experiment.	152
E.1	Histograms of prediction errors.	157

List of tables

2.1	Accidents involving gliders and motorgliders.	7
3.1	Products reviewed during the market study.	27
4.1	Characteristics of different factor levels of flight condition.	49
4.2	Relative orientation of coordinate systems during different levels of flight condition and display format (mathematical description). . . .	50
4.3	Relative orientation of coordinate systems during different flight conditions (graphical description).	51
4.4	Demographics and flight experience of participants.	52
4.5	Participants' familiarity with their assigned display format.	53
4.6	Results of video analysis.	57
5.1	Post-hoc analysis: Comparing the difference in transformed visual search error $\Delta\tilde{\gamma}$ between pairs of display format factor levels.	68
5.2	Post-hoc analysis: Comparing the difference in transformed visual search error $\Delta\tilde{\gamma}$ between pairs of flight condition factor levels.	69
5.3	Multiple regression results: Parameter estimates with 95% confidence intervals and structure coefficients.	71
6.1	Characteristics of the proposed mental models.	89
6.2	Coding of quasi-continuous relative bearing and elevation stimuli of the radar-style and perspective displays.	93
6.3	Absolute frequencies of how often different proposed mental models were identified as best-fitting.	95
6.4	Pre-planned contrasts: Comparing the difference in transformed visual search error $\Delta\tilde{\gamma}$ between groups of best-fitting mental models for different flight condition factor levels.	101
A.1	Characteristics of the traffic pointer coordinate system.	136
A.2	Characteristics of the horizon- and track-fixed coordinate system. . .	136
A.3	Characteristics of the ownship-fixed coordinate system.	137
A.4	Characteristics of the horizon- and heading-fixed coordinate system. .	138

C.1	Sex of participants, grouped by display format.	145
C.2	Age and flight experience of participants, grouped by display format.	146
D.1	Results of reaction times during answer sub-task experiment.	153

A Coordinate systems and transformations

If the visual lookout behavior of pilots is to be quantified, a mathematical description of traffic encounter geometries is required. In the following appendix, geometric descriptions, which analytically solve relevant parameters, are derived. Euclidean space is assumed and aircraft are reduced to points in space. Furthermore, these mathematical considerations are conducted from the point of view of a given pilot operating the *ownship* aircraft. *Traffic* is defined as all other aircraft in space, as seen from the ownship.

A.1 Coordinate systems

Initially, several coordinate systems (CSs) with pre-specified traits are defined. Each of these proposed CSs is well suited for providing specific geometric descriptions.

A.1.1 Traffic pointer coordinate system

Consider the vector \vec{X}^T extending from the ownship to the traffic's position. This vector can be analyzed in an arbitrary CS. Now consider fixing a Cartesian CS with the characteristics described in Table A.1 to \vec{X}^T . The \vec{x}_{tp} -axis of this CS always points in the direction of the traffic, as seen from the ownship. The CS is therefore assigned the name of *traffic pointer CS* and assigned the subscript $_{tp}$. The slant distance R to the traffic is easily determined by

$$\vec{X}_{tp}^T = R \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}_{tp} . \quad (\text{A.1})$$

Table A.1.: Characteristics of the traffic pointer coordinate system.

Axis	orientation
\vec{x}_{tp}	towards traffic aircraft
\vec{y}_{tp}	in horizontal plane, normal to \vec{x}_{tp} , pointing right
\vec{z}_{tp}	normal to \vec{x}_{tp} and \vec{y}_{tp} , pointing down
Origin at position of ownship	

A.1.2 Horizon- and track-fixed coordinate system

The fact that FLARM lacks an attitude and heading reference system means that all position data of traffic must be expressed in a CS which can be derived through Global Positioning System measurements. As the ownship's current position and position history is known in World Geodetic System 1984 coordinates [65, Sentence GPRMC], another CS is introduced. This Cartesian CS is solely derived from the current position and position history of the FLARM transceiver. It is referred to as the *horizon- and track-fixed CS* and carries the subscript $_{ht}$. The \vec{x}_{ht} - \vec{y}_{ht} -plane always remains horizontal while the \vec{x}_{ht} points in the direction of the ownship's ground track. The CS's characteristics are given in Table A.2.

Table A.2.: Characteristics of the horizon- and track-fixed coordinate system.

Axis	orientation
\vec{x}_{ht}	in the horizontal plane, oriented parallel to the ownship's track
\vec{y}_{ht}	in horizontal plane, normal to \vec{x}_{ht} , pointing 90° to the right of the ownship's track
\vec{z}_{ht}	pointing downwards in the direction of gravity
Origin at position of ownship	

A.1.3 Ownship-fixed coordinate system

The glider pilot's position within the ownship usually remains constant. When describing the traffic's position relative to the pilot's fixed position within the aircraft, an *ownship-fixed CS* may be appropriate. Therefore the ownship-fixed CS (subscript os) is introduced. The orientation of its axes and location of its origin are described in Table A.3. This is also the CS by suggested by common low-complexity displays available for the FLARM collision alerting system.

Table A.3.: Characteristics of the ownship-fixed coordinate system.

Axis	orientation
\vec{x}_{os}	parallel to the ownship's longitudinal axis, pointing in the primary direction of flight
\vec{y}_{os}	parallel to the ownship's lateral axis, pointing starboard
\vec{z}_{os}	parallel to the ownship's vertical axis, pointing in the ventral direction
Origin at position of ownship	

A.1.4 Horizon- and heading-fixed coordinate system

Several display formats suggest a CS where the reference plane is aligned with the horizontal plane. The suggested reference direction is in the direction of the ownship's heading. One such CS exhibiting these characteristics is the *horizon- and heading-fixed CS* (subscript hh). It is defined in Table A.4.

A.2 Coordinate transformations

Following the conventions of flight dynamics, angular relations can often be described as a series of rotations. These rotations transform a vector's components expressed in one CS into another CS. Transformations between the previously introduced CSs are shown in Figure A.1 and are derived in the following sections.

Table A.4.: Characteristics of the horizon- and heading-fixed coordinate system.

Axis	orientation
\vec{x}_{hh}	in the horizontal plane, oriented in the direction of the ownship's heading
\vec{y}_{hh}	in horizontal plane, normal to \vec{x}_{hh} , pointing 90° to the right of the ownship's heading
\vec{z}_{hh}	pointing downwards in the direction of gravity
Origin at position of ownship	

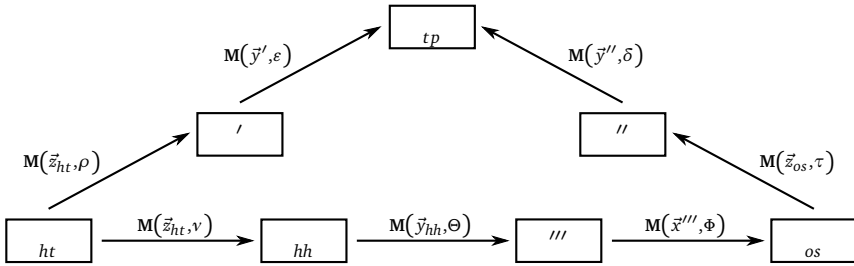


Figure A.1.: Sequence of transformation steps between coordinate systems. [illustration by author]

A.2.1 Transformation from horizon- and track-fixed coordinates to traffic pointer coordinates

Transforming a vector from horizon- and track-fixed coordinates to traffic pointer coordinates can be achieved by the following series of rotations:

1. rotation around the \vec{z}_{ht} -axis through the relative bearing $\rho \in (-180^\circ, 180^\circ]$: $\mathbf{M}(\vec{z}_{ht}, \rho)$
2. rotation around the intermediate \vec{y}' -axis through the elevation $\varepsilon \in [-90^\circ, 90^\circ]$: $\mathbf{M}(\vec{y}', \varepsilon)$

The corresponding rotation matrix $\mathbf{M}_{tp,ht} = \mathbf{M}(\vec{y}', \varepsilon) \cdot \mathbf{M}(\vec{z}_{ht}, \rho)$ is given as

$$\mathbf{M}_{tp,ht} = \begin{bmatrix} \cos \varepsilon \cos \rho & \cos \varepsilon \sin \rho & -\sin \varepsilon \\ -\sin \rho & \cos \rho & 0 \\ \sin \varepsilon \cos \rho & \sin \varepsilon \sin \rho & \cos \varepsilon \end{bmatrix}. \quad (\text{A.2})$$

The traffic pointer vector \vec{X}^T can also be expressed in horizon- and track-fixed coordinates.

$$\begin{aligned} \vec{X}_{ht}^T &= \mathbf{M}_{ht,tp} \cdot \vec{X}_{tp}^T \\ &= \mathbf{M}_{tp,ht}^{tr} \cdot \vec{X}_{tp}^T \\ &= R \cdot \begin{pmatrix} \cos \varepsilon \cos \rho \\ \cos \varepsilon \sin \rho \\ -\sin \varepsilon \end{pmatrix}_{ht} \end{aligned} \quad (\text{A.3})$$

Typically, FLARM-compatible low-complexity displays use these polar coordinates (relative bearing ρ and elevation ε) as two-dimensional indications of the traffic's positions. This coordinate transformation has a major technical advantage. Whenever the ownship's current position and position history is known, the relative bearing ρ and elevation ε to the traffic can be calculated solely on position reports of the traffic. No knowledge of the ownship's attitude is required, therefore reducing the technical complexity of the collision alerting system. However, the pilot must be aware of how the horizon- and track-fixed CS is located in order to correctly interpret the relative bearing and elevation information to the traffic.

A.2.2 Transformation from ownship-fixed coordinates to traffic pointer coordinates

Transforming a vector from ownship-fixed coordinates to traffic pointer coordinates can be achieved by the following series of rotations:

1. rotation around the \vec{z}_{os} -axis through the visual bearing $\tau \in (-180^\circ, 180^\circ]$: $\mathbf{M}(\vec{z}_{os}, \tau)$
2. rotation around the intermediate \vec{y}'' -axis through the visual elevation $\delta \in [-90^\circ, 90^\circ]$: $\mathbf{M}(\vec{y}'', \delta)$

The corresponding rotation matrix $\mathbf{M}_{tp,os} = \mathbf{M}(\vec{y}'', \delta) \cdot \mathbf{M}(\vec{z}_{os}, \tau)$ is given as

$$\mathbf{M}_{tp,os} = \begin{bmatrix} \cos \delta \cos \tau & \cos \delta \sin \tau & -\sin \delta \\ -\sin \tau & \cos \tau & 0 \\ \sin \delta \cos \tau & \sin \delta \sin \tau & \cos \delta \end{bmatrix}. \quad (\text{A.4})$$

The traffic pointer vector \vec{X}^T can also be expressed in ownship-fixed coordinates.

$$\begin{aligned} \vec{X}_{os}^T &= \mathbf{M}_{os,tp} \cdot \vec{X}_{tp}^T \\ &= \mathbf{M}_{tp,os}^{tr} \cdot \vec{X}_{tp}^T \\ &= R \cdot \begin{pmatrix} \cos \delta \cos \tau \\ \cos \delta \sin \tau \\ -\sin \delta \end{pmatrix}_{os} \end{aligned} \quad (\text{A.5})$$

A set of visual bearing τ and visual elevation δ refers to a specific spot in the out-the-window view. Therefore traffic exhibiting constant visual bearing τ and visual elevation δ will remain on a constant position on the ownship's windscreen or windows.

A.2.3 Transformation from horizon- and track-fixed coordinates to horizon- and heading-fixed coordinates

Transforming a vector from horizon- and track-fixed coordinates to horizon- and heading-fixed coordinates can be achieved by a single rotation.

1. rotation around the \vec{z}_{ht} -axis through the drift angle $\nu \in [-180^\circ, 180^\circ]$:
 $\mathbf{M}(\vec{z}_{ht}, \nu)$

The corresponding rotation matrix $\mathbf{M}_{hh,ht} = \mathbf{M}(\vec{z}_{ht}, \nu)$ is given as

$$\mathbf{M}_{hh,ht} = \begin{bmatrix} \cos \nu & \sin \nu & 0 \\ -\sin \nu & \cos \nu & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (\text{A.6})$$

A.2.4 Transformation from horizon- and track-fixed coordinates to ownship-fixed coordinates

Glider pilots are exposed to the concepts of pitch, bank and drift angles early on in the pre-solo phase in their flight training [7, chap. 1; 58, chap. 3]. Therefore these angles should provide intuitive description of the ownship's attitude to the pilot.

Transforming a vector from horizon- and track-fixed coordinates to ownship-fixed coordinates can be achieved by the following series of rotations:

1. rotation around the \vec{z}_{ht} -axis through the drift angle $\nu \in (-180^\circ, 180^\circ]$: $\mathbf{M}(\vec{z}_{ht}, \nu)$
2. rotation around the \vec{y}_{hh} -axis through the pitch angle $\Theta \in [-90^\circ, 90^\circ]$: $\mathbf{M}(\vec{y}_{hh}, \Theta)$
3. rotation around the intermediate \vec{x}''' -axis through the bank angle $\Phi \in (-180^\circ, 180^\circ]$: $\mathbf{M}(\vec{x}''', \Phi)$

The corresponding rotation matrix $\mathbf{M}_{os,ht} = \mathbf{M}(\vec{x}''', \Phi) \cdot \mathbf{M}(\vec{y}_{hh}, \Theta) \cdot \mathbf{M}(\vec{z}_{ht}, \nu)$ is given as

$$\mathbf{M}_{os,ht} = \begin{bmatrix} \cos \Theta \cos \nu & & & \\ -\cos \Phi \sin \nu + \sin \Phi \sin \Theta \cos \nu & & & \\ \sin \Phi \sin \nu + \cos \Phi \sin \Theta \cos \nu & \dots & & \\ & \cos \Theta \sin \nu & & -\sin \Theta \\ & \cos \Phi \cos \nu + \sin \Phi \sin \Theta \sin \nu & & \sin \Phi \cos \Theta \\ \dots & -\sin \Phi \cos \nu + \cos \Phi \sin \Theta \sin \nu & & \cos \Phi \cos \Theta \end{bmatrix}. \quad (\text{A.7})$$

A.3 Describing differences between actual and suspected positions of traffic

As previously described, the vector \vec{X}^T represents the direction to the traffic relative to the ownship's location. However, a pilot may suspect the traffic to be at a different location. In this case, the location where the pilot suspects traffic to be is described by the vector \vec{X}^S . This vector can be expressed in ownship-fixed coordinates using the pilot's suspected slant distance R^S , visual elevation δ^S and visual

bearing τ^S . These three values are defined analogously to their actual counterparts, allowing the vector \vec{X}^S to be expressed as

$$\vec{X}_{os}^T = R^S \cdot \begin{pmatrix} \cos \delta^S \cos \tau^S \\ \cos \delta^S \sin \tau^S \\ -\sin \delta^S \end{pmatrix}_{os}. \quad (A.8)$$

ΔR , $\Delta \delta$ and $\Delta \tau$ represent the errors in slant distance, visual elevation and visual bearing between the traffic's actual and suspected positions.

$$\Delta R = R^S - R \quad (A.9a)$$

$$\Delta \delta = \delta^S - \delta \quad (A.9b)$$

$$\Delta \tau = \tau^S - \tau \quad (A.9c)$$

The visual search error $\Delta \gamma$ is defined as the intersecting angle between the actual and suspected position vectors.

$$\Delta \gamma = \angle (\vec{X}^T, \vec{X}^S) \quad (A.10)$$

Solving equation A.10 is a standard problem of analytical geometry [22, chap. 3].

$$\begin{aligned} \cos \Delta \gamma &= \frac{\vec{X}^T \cdot \vec{X}^S}{|\vec{X}^T| \cdot |\vec{X}^S|} \\ &= \frac{R \cdot \begin{pmatrix} \cos \delta \cos \tau \\ \cos \delta \sin \tau \\ -\sin \delta \end{pmatrix}_{os} \cdot R^S \cdot \begin{pmatrix} \cos \delta^S \cos \tau^S \\ \cos \delta^S \sin \tau^S \\ -\sin \delta^S \end{pmatrix}_{os}}{R \cdot R^S} \\ &= \cos \delta \cdot \cos \delta^S \cdot [\sin \tau \cdot \sin \tau^S + \cos \tau \cdot \cos \tau^S] + \sin \tau \cdot \sin \tau^S \\ &= \frac{1}{2} \left[[\cos \Delta \tau + 1] \cos \Delta \delta + [\cos \Delta \tau - 1] \cos (2\delta^S - \Delta \delta) \right] \end{aligned} \quad (A.11)$$

The visual search error magnitude $\Delta \gamma$ therefore is

$$\Delta \gamma = \cos^{-1} \left(\frac{1}{2} \left[[\cos \Delta \tau + 1] \cos \Delta \delta + [\cos \Delta \tau - 1] \cos (2\delta^S - \Delta \delta) \right] \right), \quad (A.12)$$

with $\Delta \gamma \in [0^\circ, 180^\circ]$.

B Interpretation intervals for effect sizes

Estimated effect sizes of statistically significant results reveal the strength of relationships [4, pp. 25–26]. They aid in extending the concept of statistical significance to substantive significance [128]. Interpreting the numerical values of effect sizes requires interpretation intervals to be defined. However, B. H. Cohen noted that effect sizes should, whenever possible, be seen in context of experience with similar experiments [31, pp. 244–245]. Therefore, the following intervals form no hard boundaries but should be used as a general reference for interpreting the results.

B.1 Cramér’s V and Pearson’s bivariate r

Cramér’s V is an effect size used during tests of cross tables. Similarly, Pearson’s r describes the strength of bivariate correlations between variables. B. H. Cohen defined interpretation intervals for Cramér’s V [31, pp. 727–728], while J. Cohen named such intervals for Pearson’s r [32, section 3.2]. The interval boundaries were defined identically in both cases as follows:

0	≤	$V, r $	<	.1,	negligible effect
.1	≤	$V, r $	<	.3,	small effect
.3	≤	$V, r $	<	.5,	moderate effect
.5	≤	$V, r $,	large effect

B.2 Pearson’s \bar{R}^2

Bortz stated that Pearson’s multivariate regression coefficient \bar{R}^2 (*adjusted R^2*) should be interpreted similarly to the bivariate coefficient [19, pp. 449–451]. Therefore, it is interpreted using the following intervals.

0	≤	\bar{R}^2	<	.01,	negligible effect
.01	≤	\bar{R}^2	<	.09,	small effect
.09	≤	\bar{R}^2	<	.25,	moderate effect
.25	≤	\bar{R}^2		,	large effect

B.3 η^2 and generalized η^2

J. Cohen recommended interpretation intervals for the η^2 effect size [32, pp. 413–414], which is generally used during analyses of variance. Bakeman proposed that the generalized η^2 effect sizes (η_G^2) for repeated measures designs should be interpreted using the same intervals [11].

0	\leq	η^2, η_G^2	$< .02,$	negligible effect
.02	\leq	η^2, η_G^2	$< .13,$	small effect
.13	\leq	η^2, η_G^2	$< .26,$	moderate effect
.26	\leq	η^2, η_G^2	,	large effect

C Variations of flight experience and demography with display format

Participants were assigned to one of three display formats, based on the date that they participated in the study. This procedure held the potential that differences in the participants’ demography and glider flight experience might skew the results. No a priori indications of such skewing were recognized. Nevertheless, to confirm that no skewing took place, the following hypothesis was evaluated.

Hypothesis 6: The participants’ age, sex and glider flight experience do not vary with the assigned display.

Demographic data and flight experience of the participants according to the display assigned are shown in Tables C.1 and C.2.

Table C.1.: Sex of participants, grouped by display format.

Display format	female	male
low-complexity	3	39
radar-style	4	29
perspective	1	45

Two statistical tests were performed to evaluate whether participants’ age, sex and glider flight experience varied between display formats. The first test was a Fisher’s exact test and the second test was multivariate analysis of variance (MANOVA). In both tests the between-subjects independent variable was defined as the display format assigned to participants with the factor levels of *low-complexity*, *radar-style* and *perspective* displays. Since Hypothesis 6 was tested at $\alpha = .05$, the significance level of each of the two tests was adjusted using Holm’s step down procedure.

Variations of participants’ sex (nominally scaled dependent variable) and their assigned display were examined using Fisher’s exact test. The test was applied to the data of Table C.1. It provided statistically insignificant results at the adjusted significance level ($\chi^2(2) = 3.08, p = .21$).

Table C.2.: Age and flight experience of participants, grouped by display format (between-subjects mean M and standard deviation (SD)).

Parameter	display format		
	low-complexity	radar-style	perspective
Age in years	25.8 (10.9)	25.0 (9.6)	31.3 (17.6)
Total flight time in gliders in hr	222.6 (192.4)	290.9 (694.1)	409.2 (714.2)
Glider flight time within last 6 months in hr	13.3 (15.1)	8.9 (12.0)	17.5 (27.5)
Total number of glider flights	441.6 (427.7)	370.6 (422.5)	641.1 (936.0)

A MANOVA was used to evaluate whether the interval scaled dependent variables of demography and experience varied between display formats. Due to high correlations between age and most of the flight experience metrics of Table C.2, only age and glider flight time within the preceding six months were retained as the MANOVA's dependent variables. A robust nonparametric variant of the MANOVA [141] was performed. Equality of the covariance matrices was given.¹ The MANOVA's results were statistically insignificant at the adjusted significance level as well (test score $L(2) = 4.47$, $p = .11$).

These results of the MANOVA and Fisher's exact test were congruent with Hypothesis 6. Thus, it was retained. No systematic difference in demography or flight experience existed between participants assigned to different display formats.

¹ analyzed using Box's M test



D Experimentally evaluating the answer sub-task

The experimental task of this thesis's main experiment, which is discussed in section 4.5, consisted of three sub-tasks. At first, participants were to notice and interpret the traffic signal shown to them. During the second sub-task, participants directed their gaze to the out-the-window view. During the final sub-task, they tapped an answer cell on a touch-screen monitor to indicate where they previously looked. The answer sub-task is not found in actual glider flight operations. Instead, it was inherent only to the main experiment.

D.1 Introduction

While studying the results relevant to actual flight operations, it was necessary to compensate for the influence of the answer sub-task. This required two categories of questions about the answer sub-task to be discussed.

- How often do participants provide the touch screen monitor response they intend to give? How high is the error rate when pressing the touch screen monitor's buttons?
- How long does it take the participants to provide their intended touch screen monitor response? What influences their response time to the answer sub-task?

During the design phase of the main experiment, several research pilots from Technische Universität Darmstadt's flight department provided their expert opinions on the tasks. From these expert opinions two hypotheses regarding the answer sub-task were derived.

Hypothesis 7: The success rates during the answer sub-task are adequate for the experiment. Participants almost exclusively press the buttons they intend to press.

Hypothesis 8: Response time varies between cells.

D.2 Method

In order to test the aforementioned hypotheses, an answer sub-task experiment was designed. In this experiment, the main experiment's answer sub-task (Figures 4.7 and 4.8) was performed as a stand-alone task by a new sample of participants. This allowed insight into the characteristics of the answer sub-task.

A total of 28 participants performed the answer sub-task experiment. 3 participants were female, the remainder male. Their ages ranged between 24 and 51 years of age ($M = 28.2$ years, $SD = 5.4$ years). All participants were either students of Technische Universität Darmstadt or scientific personnel affiliated with the Institute of Flight Systems and Automatic Control. They did not require flight experience to participate.

Participants performed the answer sub-task experiment in a quiet office environment. They seated themselves at a conventional office workplace. There, they took hold of the touch screen monitor also used during the main experiment. The touch screen monitor showed the main experiment's answer grid consisting of 50 cells (Figure 4.2). During the answer sub-task experiment, the alphanumeric identifiers of all cells were displayed in randomized order in 4 s intervals on a laptop computer screen. Participants were instructed to perform the following task:

- a) Look at the laptop computer screen. As soon as you notice a new alphanumeric identifier, read the identifier.
- b) Swiftly locate and tap the corresponding cell on the answer grid shown on the touch screen monitor.

The button pressed by the participant as well as the corresponding response time were recorded as the answer sub-task experiment's dependent variables. Independent variables were the row and column identifiers of the cell, as shown on the laptop computer screen. The hardware setup, shown in Figure D.1, was a simplified setup of main experiment's setup (see section 4.3). Similarly, a modified version of the experimental software was developed for this purpose [100, section 5.5.2]. Since the experiment was performed by multiple participants, the participant's running identification number was as a repeated measures independent variable. This resulted in a $5 \times 10 \times 28$ within-subject repeated measures design of the experiment.

By comparing the alphanumeric identifiers shown to the participants to the cells they pressed, each participant's success rate in pressing the correct button on the touch screen monitor was calculated. A success rate of 90% or above was arbitrarily defined as being adequate within the frame of this experiment. Also, a limited error success rate was determined for each participant. It described the relative number

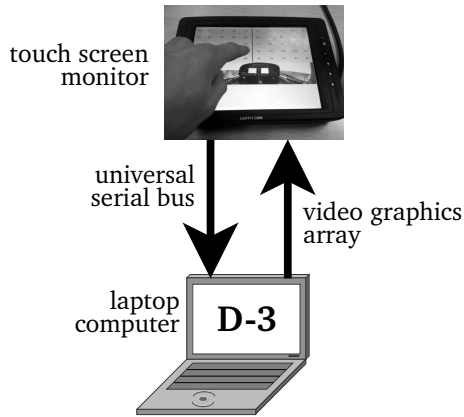


Figure D.1.: Hardware setup of the answer sub-task experiment. [illustration and photograph by author]

of responses where a participant tapped a response cell which was within one row vertically and one column horizontally of the commanded cell. Also, response time was evaluated. Only the response time of correct responses was included. Outliers were removed.

D.3 Results

Determining whether participants' responses were adequate (Hypothesis 7) was performed by descriptively analyzing their actual (Figure D.2a) and limited error (Figure D.2b) success rates. The actual success rate distribution had a between-subjects mean $M = 96.4\%$ ($SD = 3.9\%$). Its 10th percentile was located at a success rate of 90.6%. Naturally, the limited error success rate's between-subjects mean was higher at $M = 98.2\%$ ($SD = 3.3\%$). The corresponding 10th percentile was also higher at 96.0%. All participants having scored a success rate of less than 90% during the experiment lacked answers for more than one out of the 50 signals shown. Inversely, all participants having scored a success rate of more than 90% had only one or no missing values.

Influences on the participants' response time (Hypothesis 8) were determined using inferential statistics. The participants' response time was grouped by the

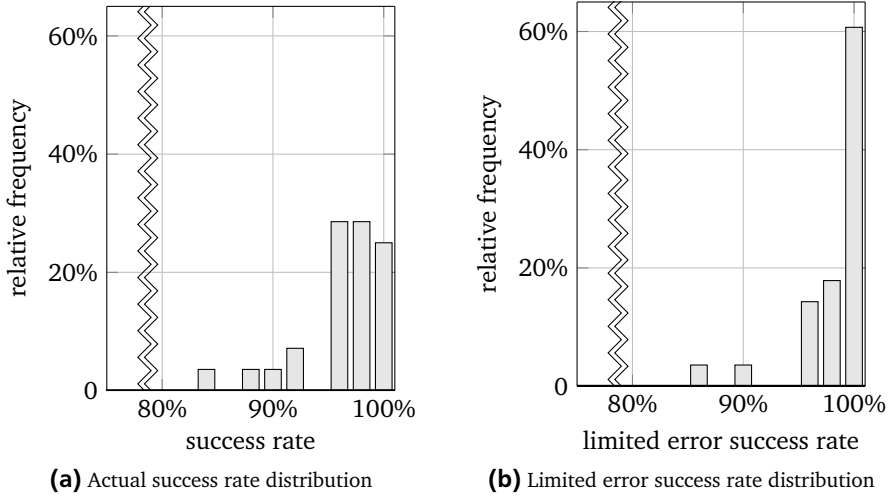


Figure D.2.: Experimental distributions of actual and limited error success rates during answer sub-task experiment.

corresponding answer cell. All distributions of response time were normal¹ and exhibited heteroscedasticity.² The between-subjects means and standard deviations for each cell are shown in Table D.1.

Hypothesis 8 was analyzed using a two-way analysis of variance. Its dependent variable was the response time of each participant. The column identifiers (−5 through −1 and 1 through 5) and row identifiers (A through E) - as answered by the participants - formed the analysis of variance's two independent variables. Due to the data's heteroscedasticity, the robust F^* test score proposed by Brown and Forsythe was used [24]. Both independent variables as well as their interaction had statistically significant effects on the participants' reaction times ($F^*(4, 479.60) = 17.46, p < .01, \eta^2 = .05$ for row effect; $F^*(9, 479.60) = 2.08, p = .03, \eta^2 = .01$ for column effect; $F^*(36, 479.60) = 2.87, p < .01, \eta^2 = .07$ for interaction between rows and columns). While the effect of the column on reaction time was of negligible strength, the row and interacting effects were small, but notable.

¹ assessed using 50 Kolmogorov-Smirnov tests for normality. All tests provided statistically insignificant results at the Holm's step-down adjusted significance level.

² assessed using the Levene test for homogeneity of variances. Its results were statistically significant with $F(49, 1297) = 2.03, p < .01$.

Table D.1.: Results of reaction times during answer sub-task experiment.

(a) Between-subjects mean of reaction time in s										
Row	Column									
	-5	-4	-3	-2	-1	1	2	3	4	5
A	1.88	1.81	1.55	1.48	1.30	1.27	1.38	1.48	1.61	1.60
B	1.88	1.90	1.82	1.72	1.84	1.66	1.42	1.70	1.78	1.78
C	1.84	1.89	1.76	1.77	1.85	1.59	1.69	1.61	1.73	1.96
D	1.64	1.72	1.85	1.93	1.93	1.88	1.91	1.75	1.63	2.03
E	1.66	1.74	1.96	1.72	1.83	1.81	1.89	1.97	1.98	1.61

(b) Between-subjects standard deviation of reaction time in s										
Row	Column									
	-5	-4	-3	-2	-1	1	2	3	4	5
A	0.53	0.41	0.34	0.34	0.30	0.25	0.25	0.33	0.31	0.34
B	0.54	0.39	0.50	0.51	0.81	0.48	0.36	0.45	0.36	0.39
C	0.45	0.47	0.46	0.52	0.62	0.54	0.45	0.29	0.33	0.44
D	0.49	0.44	0.57	0.41	0.44	0.44	0.41	0.38	0.39	0.44
E	0.44	0.58	0.56	0.55	0.60	0.49	0.48	0.55	0.60	0.39

D.4 Discussion

The analysis of the success rate showed that for almost all participants 90% or more of their answers were precise. In these cases, participants pressed the answer cells on the touch screen monitor which they intended to press. Whenever an error occurred, that error was almost exclusively within one row and one column of the intended answer cell on the monitor's answer grid. Those 2 of 28 participants having scored success rates of less than 90% had problems performing their experimental tasks and therefore missed to perform their task properly on several occasions. The scores of these participants were considered to be outliers. The fact that 90% of the participants scored actual success rates of 90.6% or higher, and limited error success rates of 96.0% or better, illustrates that the answer sub-task experiment was performed with acceptable performance by the participants. It is assumed that this acceptable performance carries over into the main experiment. Hypothesis 7 was therefore retained on the basis of these descriptive results.

Results of the analysis of variance showed that the participants' response time varied notably between individual answer cells. Accordingly, Hypothesis 8 was retained. Due to the non-negligible interaction between rows and columns on response time, adequate corrections should be performed on the reaction time measured during the main experiment. One way of doing so is to subtract the between-subjects mean response time of Table D.1a from the main experiment's measured reaction time (Figure 4.8). Additionally, this allows for a more plausible definition of outliers in the main experiment (section 4.6).

E Verifying the analytical modeling approach to identifying pilots' mental models

Chapter 6 presented a novel approach to identifying the mental models (MMs) which glider pilots use when interacting with a low-cost collision alerting system (CAS). In this approach, MMs of glider pilots were modeled analytically. This was achieved by proposing multiple coordinate systems (CSs) to predict pilot responses to signals shown on the displays of their CAS. The CS - and corresponding MM - best predicting the participant's expectations where traffic would be located was then selected as the pilot's best-fitting MM. Before using this novel approach it had to be checked for plausibility and verified.

E.1 Introduction

For verification, the analytical modeling approach used to identify MMs had to provide plausible results. For no factor level of flight condition should the analytical modeling approach have skewed the predicted direction of the participants' answers to their task, when compared to their actual answer. Also, the best-fitting MMs should correspond to a notable variation in the participants' prediction error magnitude $\Delta\gamma^P$. The initial motivation behind analyzing the MMs was to explain differences in pilot effectivity when using the CAS. If the analytical modeling approach can't explain these results, then it is considered inappropriate for the designed task. These considerations led to two separate validation hypotheses.

Hypothesis 9: Throughout all flight conditions, the best-fitting MM for each participant does not skew prediction results. Instead, answers predicted by the best-fitting MM scatter symmetrically around the participant's actual answers.

Hypothesis 10: The best-fitting MM influences the effectivity demonstrated by participants using a low-cost CAS.

E.2 Method

Identification of the best-fitting MM for each participant was performed according to section 6.2. For verification purposes not only the magnitude $\Delta\gamma^P$ of the prediction error was of interest but also the direction in which the prediction error occurred. By comparing the predicted visual bearing τ^P of the analytical modeling approach against the suspected visual bearing τ^S of the participant, the visual bearing prediction error $\Delta\tau^P$ was defined. Analogously, a visual elevation prediction error $\Delta\delta^P$ was defined as well. Both prediction errors decomposed the prediction error's magnitude $\Delta\gamma^P$ into two independent directions.

$$\Delta\tau^P = \tau^P - \tau^S \quad (\text{E.1a})$$

$$\Delta\delta^P = \delta^P - \delta^S \quad (\text{E.1b})$$

Outliers were removed, and where necessary, they were conservatively replaced with the participant's average values of the respective dependent variable for that factor level of flight condition. The assumptions and prerequisites necessary for the following statistical procedures were checked beforehand. Where necessary, corrections were performed.

E.3 Results and discussion

Hypothesis 9 was examined using multiple Wilcoxon signed rank tests. For each factor level combination of participant and flight condition, the distributions of visual bearing prediction error $\Delta\tau^P$ and visual elevation prediction error $\Delta\delta^P$ were determined (see equations E.1). Their frequency distributions are shown in Figure E.1. Each of these distributions was tested whether it conformed to a symmetric distribution around the zero median using a one-sample Wilcoxon signed rank test. This resulted in eight tests per participant and 968 tests in total. Of these 968 tests, 18 tests provided statistically significant results at a Holm's step-down adjusted significance level. However, all corresponding rank correlation coefficients r of these statistically significant tests revealed a negligible effect size; $|r| < .08$. The results revealed that for each factor level combination of participant and flight condition, the best-fitting MM did not cause substantive systematic errors in predicting the participants' responses. Instead, the predictions scattered symmetrically around the participants' actual answers in the visual bearing direction and visual elevation direction. Thus, Hypothesis 9 was retained.

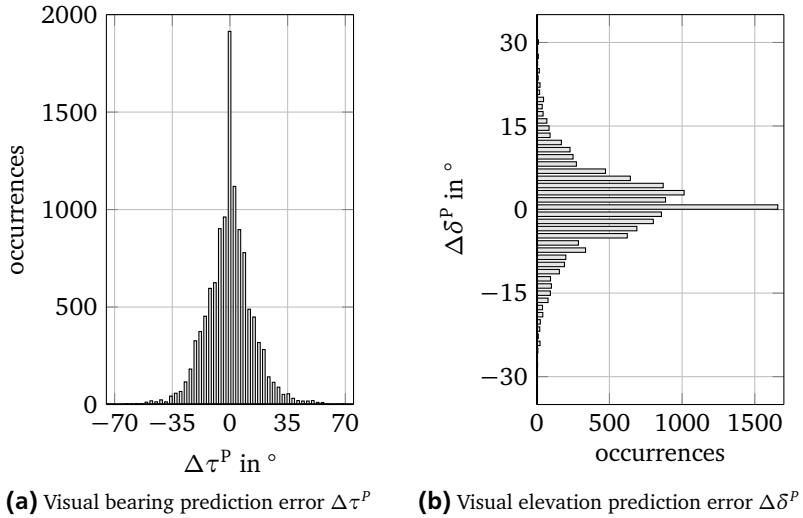


Figure E.1.: Histograms of prediction errors.

Testing the influence of the best-fitting MM on the participants' effectivity in using the CAS (Hypothesis 10) was achieved by using a four-way mixed-design multivariate analysis of variance (MANOVA). The MANOVA's between-subjects independent variables were the best-fitting MM and the display format. Its within-subject independent variables were the flight condition and the signal number. The transformed visual search error $\Delta\tilde{\gamma}$ served as the MANOVA's only dependent variable. In order to reduce the model's degrees of freedom, only main effects and limited interacting effects were modeled. These interacting effects were the 'display format * flight condition', 'best-fitting MM * flight condition', 'display format * signal number' and 'best-fitting MM * signal number' interactions. As discussed in section 5.1, the selected MANOVA structure was moderately robust against violations of the normality and homoscedasticity assumptions. Therefore, these assumptions were not assessed further. The main effect of the best-fitting MM on $\Delta\tilde{\gamma}$ was statistically significant; $F(7,111) = 3.20$, $p < .01$. The corresponding effect size $\eta_G^2 = .03$ revealed a small effect. This showed that the analytical modeling approach could be used to explain differences in pilot effectivity. Hypothesis 10 was retained.

E.4 Conclusion

In summary, the analytical modeling approach used to identify the MMs of glider pilots behaved plausibly. No systematic errors in predicting the responses of participants were made. The approach fulfilled its design goal by uncovering performance differences exhibited by participants using a low-cost CAS. Since none of the results negated the approach presented, verification of the mental modeling approach was considered to be successful.